December 1954

BIOMETRICS

Vol. 10 No. 4

JOURNAL OF THE BIOMETRIC SOCIETY

Some Methods for Strengthening the Common χ^2 Tests
W. G. Cochran

Short-Cut Formulas for the Exact Partition of χ^2 in Contingency Tables — *A. W. Kimball*

A Note on the χ^2 Test Applied to Epidemic Chains
Joyce Almond

Misclassification in 2 x 2 Tables — *Irwin Bross*

A 2 x 2 Factorial with Paired Comparisons
Robert M. Abelson and Ralph A. Bradley

Some Considerations in the Design of Successive Experiments in Fruit Plantations
T. N. Hoblyn, S. C. Pearce and G. H. Freeman

Variance Heterogeneity in a Randomized Block Design
Franklin Graybill

Some Observations on Sensitivity and Repeatability of Triad Taste Difference Tests — *J. W. Hopkins*

The Estimation of Response-Time Distributions III Truncation and Survival — *M. R. Sampford*

The Biometric Society

BIOMETRICS

FOUNDED BY THE BIOMETRICS SECTION OF THE AMERICAN STATISTICAL ASSOCIATION

TABLE OF CONTENTS

Some Methods for Strengthening the Common χ^2 Tests
W. G. Cochran 417

Short-Cut Formulas for the Exact Partition of χ^2 in Contingency Tables . . . . . A. W. Kimball 452

A Note on the χ^2 Test Applied to Epidemic Chains
Joyce Almond 459

Misclassification in 2 × 2 Tables . . . . . Irwin Bross 478

A 2 × 2 Factorial with Paired Comparisons
Robert M. Abelson and Ralph A. Bradley 487

Some Considerations in the Design of Successive Experiments in Fruit Plantations
T. N. Hoblyn, S. C. Pearce and G. H. Freeman 503

Variance Heterogeneity in a Randomized Block Design
Franklin Graybill 516

Some Observations on Sensitivity and Repeatability of Triad Taste Difference Tests . . . . . J. W. Hopkins 521

The Estimation of Response-Time Distributions. III Truncation and Survival . . . . . M. R. Sampford 531

Queries 562

Abstracts 569

The Biometric Society 571

Table of Contents to Vol. 10 and Index to Vols. 1–10 577

Number 4 December 1954 Volume 10

Material for *Biometrics* should be addressed to Miss Gertrude Cox, Institute of Statistics, Box 5457, Raleigh, North Carolina, except that authors residing in one of the following organized regions can expedite the handling of their papers by submitting them to the Assistant Editor for that region.

British Region: Dr. D. J. Finney, Dept. of Stat., Univ. of Aberdeen, Aberdeen, Scotland; **Australasian Region**: Dr. E. A. Cornish, University of Adelaide, Adelaide, Australia; **French Region**: Dr. Georges Teissier, Faculte des Sciences de Paris, 1 rue V. Cousin, Paris, France.

Material for Queries should go to Professor G. W. Snedecor, Statistical Laboratory, Iowa State College, Ames, Iowa.

Articles to be considered for publication should be submitted in triplicate.

THE BIOMETRIC SOCIETY

General Officers

President, W. G. Cochran; *Secretary-Treasurer*, C. I. Bliss; *Council*, H. C. Batson, L. L. Cavalli-Sforza, Georges Darmois, C. W. Emmens, D. J. Finney, Sir Ronald Fisher, J. O. Irwin, Arthur Linder, P. C. Mahalanobis, Donald Mainland, Leopold Martin, A. M. Mood, C. R. Rao, Georges Teissier, J. W. Tukey, Frank Yates, W. J. Youden.

Regional Officers

Eastern North American Region: *Regional President*, S. L. Crump; *Secretary-Treasurer*, A. M. Dutton. British Region: *Regional President*, R. R. Race; *Secretary*, E. C. Fieller; *Treasurer*, A. R. G. Owen. Western North American Region: *Regional President*, D. G. Chapman; *Secretary-Treasurer*, Elizabeth Vaughan. Australasian Region: *Regional President*, Helen N. Turner; *Secretary*, W. B. Hall; *Treasurer*, Mary A. Whitehead. French Region: *Regional President*, Georges Darmois; *Secretary Treasurer*, Daniel Schwartz. Belgian Region: *Regional President*, Paul Spehl; *Secretary*, Leopold Martin; *Treasurer*, Claude Panier. Italian Region: *Regional President*, C. Barigozzi; *Secretary*, L. L. Cavalli-Sforza; *Treasurer*, R. Scossiroli.

National Secretaries

Denmark, N. F. Gjeddebaek; The Netherlands, E. van der Laan; India, V. G. Panse; Germany, Maria-Pia Geppert; Japan, M. Hatamura; Switzerland, Arthur Linder; Sweden, H. O. A. Wold; Brazil, Americo Groszmann.

Editorial Board

Biometrics

Editor: Gertrude M. Cox; *Assistant Editors and Committee Members:* C. I. Bliss, Irwin Bross, E. A. Cornish, W. J. Dixon, Mary Elveback, Ralph Bradley, D. J. Finney, S. Lee Crump, Leopold Martin, K. R. Nair, Horace W. Norton, H. Fairfield Smith, G. W. Snedecor and Georges Teissier. *Managing Editor:* Sarah P. Carroll.

The Biometric Society is an international society devoted to the mathematical and statistical aspects of biology and welcomes to membership biologists, mathematicians, statisticians and others who are interested in its objectives. Through its regional organizations the Society sponsors regional and local meetings. National secretaries serve the interest of members in Denmark, the Netherlands, India, Germany, Japan, Sweden and Brazil and there are many members "at large". Dues in the Society for 1954 for residents of the Western Hemisphere are as follows: Full membership including subscription to *Biometrics* is $7.00. Members of the Biometrics Section of the American Statistical Association who subscribe to the journal through that organization may become members of The Biometric Society on the payment of $3.00 annual dues. For members in other parts of the world, full membership including subscription to *Biometrics* is $4.50, except that members who subscribe to the journal through the American Statistical Association pay annual dues of $1.75. Information concerning the Society can be obtained from the Secretary, The Biometric Society, Drawer 1106, New Haven 4, Connecticut, U.S.A.

Annual subscription rates to non-members are as follows: For American Statistical Association Members, $4.00; for subscribers, non-members of either American Statistical Association or The Biometric Society, $7.00. Subscriptions should be sent to the Managing Editor, *Biometrics*, P. O. Box 5457, Raleigh, North Carolina, U.S.A.

Entered as second-class matter at the Post Office at New Haven, Conn., under the Act of March 3, 1879. Additional entry at Richmond, Va. Business Office, 52 Hillhouse Ave., New Haven, Conn. *Biometrics* is published quarterly—in March June, September and December.

SOME METHODS FOR STRENGTHENING THE COMMON χ^2 TESTS*

WILLIAM G. COCHRAN
The Johns Hopkins University

1. INTRODUCTION

Since the χ^2 tests of goodness of fit and of association in contingency tables are presented in many courses on statistical methods for beginners in the subject, it is not surprising that χ^2 has become one of the most commonly-used techniques, even by scientists who profess only a smattering of knowledge of statistics. It is also not surprising that the technique is sometimes misused, e.g. by calculating χ^2 from data that are not frequencies or by errors in counting the number of degrees of freedom. A good catalogue of mistakes of this kind has been given by Lewis and Burke (1).

In this paper I want to discuss two kinds of failure to make the best use of χ^2 tests which I have observed from time to time in reading reports of biological research. The first arises because χ^2 tests, as has often been pointed out, are not directed against any specific alternative to the null hypothesis. In the computation of χ^2, the deviations $(f_i - m_i)$ between observed and expected frequencies are squared, divided by m_i in order to equalize the variances (approximately), and added. No attempt is made to detect any particular pattern of deviations $(f_i - m_i)$ that may hold if the null hypothesis is false. One consequence is that the usual χ^2 tests are often insensitive, and do not indicate significant results when the null hypothesis is actually false. Some forethought about the kind of alternative hypothesis that is likely to hold may lead to alternative tests that are more powerful and appropriate. Further, when the ordinary χ^2 test does give a significant result, it does not direct attention to the way in which the null hypothesis disagrees with the data, although the pattern of deviations may be informative and suggestive for future research. The remedy here is to supplement the ordinary test by additional tests that help to reveal the significant type of deviation.

In this paper a number of methods for strengthening or supplementing the most common uses of the ordinary χ^2 test will be presented and illustrated by numerical examples. The principal devices are as follows:

*Work assisted by a contract with the Office of Naval Research, U. S. Navy Department. Dept. of Biostatistics paper no. 278.

1. Use of small expectations in computing χ^2.
2. Use of a single degree of freedom, or a group of degrees of freedom, from the total χ^2.
3. Use of alternative tests.

Most of the techniques have been available in the literature for some time: indeed, most of them stem from early editions of Fisher's "Statistical Methods for Research Workers." Research which has clarified the problem of subdividing χ^2 in contingency tables is more recent, and still continues.

In the hope of avoiding some confusion, the symbol X^2 will be used for the quantity that we calculate from the sample in a chi-square test. The symbol χ^2 itself will refer to a random variate that follows the distribution in the χ^2 tables, and will sometimes be used in the phrase "the χ^2 test."

2. USE OF SMALL EXPECTATIONS IN COMPUTING X^2

In order to prove that the quantity $X^2 = \sum(f_i - m_i)^2/m_i$ is distributed as χ^2 when the null hypothesis is true, it is necessary to postulate that the expectations m_i are large: in fact, the proof is strictly valid only as a limiting result when the m_i tend to infinity. For this reason many writers recommend that the m_i be not less than 5 when applying the test in practice, and that neighboring classes be combined if this requirement is not met in the original data. Some writers recommend a lower limit of 10 for the m_i .

It is my opinion that these recommendations are too conservative, and that their application may on occasion result in a substantial loss of power in the test. I give this as an opinion, because not enough research has been done to make the situation quite clear. However, the exact distribution of $\sum(f_i - m_i)^2/m_i$, when the expectations are small, has been worked out in a number of particular cases by Sukhatme (2), Neyman and Pearson (3) and Cochran (4), (5). These results indicate that the χ^2 tables give an adequate approximation to the exact distribution even when some m_i are much lower than 5.

Loss of power from following a rule that $m_i \geq 5$ occurs because this rule tends to require grouping of classes at the tails or extremes of the distribution. These are often the places where the difference between the alternative hypothesis and the null hypothesis stands out most clearly, so that the grouping may cover up the most marked difference between the two hypotheses. Information about the extent of the loss of power is unfortunately very scanty, because the power function of X^2 is known only as a limiting result when the m_i are large. The following illustration suggests that the loss can be large.

Suppose that we have a sample of $N = 100$. The null hypothesis is that the data follow a Poisson distribution with mean m known to be 1, when actually the data follow the negative binomial distribution $(q - p)^{-n}$, where $n = 2$, $q = 1.5$, $p = 0.5$. What is the chance of rejecting the null hypothesis at the 5% level of significance?

The expected frequencies of 0, 1, 2, 3, and 4 or more occurrences on the two hypotheses are shown in table 1.

TABLE 1.
EXPECTED FREQUENCIES ON THE NULL AND ALTERNATIVE HYPOTHESIS.

	Expected frequencies		$\frac{(m'_i - m_i)^2}{m_i}$		
				Grouped	
i	Poisson m_i	Neg. Bin. m'_i	Ungrouped	$m_i \geq 5$	$m_i \geq 10$
0	36.79	44.44	1.590	1.590	1.590
1	36.79	29.63	1.393	1.393	1.393
2	18.39	14.82	0.693	0.693	
3	6.13	6.58	0.033	1.181	0.009
4+	1.90	4.53	3.640		
Totals N =	100.00	100.00	$\lambda = 7.349$	$\lambda = 4.857$	$\lambda = 2.992$

If an m_i as low as 1.90 is allowed, we can use 5 classes in the χ^2 test, with 4 degrees of freedom since the m_i are known. To make all $m_i \geq 5$, we must pool the last two classes and have 3 degrees of freedom; and to make all $m_i \geq 10$, we must pool the last three classes and have 2 degrees of freedom. In order to obtain an approximation to the powers of these three ways of applying the χ^2 test, we shall use the asymptotic result for the power function. This is a non-central χ^2 distribution, with parameter λ of non-centrality, where

$$\lambda = \sum \frac{(m'_i - m_i)^2}{m_i}$$

The larger the value of λ, the higher is the power. The contributions to λ from each class are shown in the right-hand columns of table 1. For the ungrouped case, note that the extreme class 4+ is much the largest contributor to λ. Grouping the last two classes considerably reduces this contribution, while grouping the last three classes diminishes it almost to zero. The approximate probabilities of rejecting the null

hypothesis may be read from Fix's tables (6) of the non-central χ^2 distribution. These probabilities are 0.56 for the ungrouped case, 0.43 for $m_i \geq 5$ and 0.32 for $m_i \geq 10$. The loss in sensitivity from grouping is evident. Perhaps a more revealing comparison is to compute from Fix's tables the sizes of sample N that would be needed in the two grouped cases in order to bring the probabilities of rejection up to 0.56, the value for the ungrouped case. The results are $N = 136$ when the last two classes are grouped and $N = 191$ when the last three classes are grouped, as against $N = 100$ in the ungrouped case.

This example is only suggestive, and probably favors the ungrouped case slightly, because the computations are based on large-sample results. However, the losses in power from grouping, measured in terms of equivalent sample sizes, are impressive.

2.1 *Recommendations about minimum expectations*

Elsewhere, (7), I have given recommendations about the minimum expectation to be used in χ^2 tests. These working rules may be summarized, in slightly revised form, as follows:

(*a*) *Goodness of fit tests of unimodal distributions* (*such as the normal or Poisson*). Here the expectations will be small only at one or both tails. Group so that the minimum expectation at each tail is at least 1.

(*b*) *The* 2×2 *table.* Use Fisher's exact test (i) if the total N of the table < 20, (ii) if $20 < N < 40$ and the smallest expectation is less than 5. Mainland (8) has given useful tables of the exact test for these cases. If $N > 40$ use X^2, corrected for continuity.

(*c*) *Contingency tables with more than* 1 *d.f.* If relatively few expectations are less than 5 (say in 1 cell out of 5 or more, or 2 cells out of 10 or more), a minimum expectation of 1 is allowable in computing X^2.

Contingency tables with most or all expectations below 5 are harder to prescribe for. With very small expectations, the exact distribution of X^2 can be calculated without too much labor. Computing methods have been given by Freeman and Halton (9). If X^2 has less than 30 degrees of freedom and the minimum expectation is 2 or more, use of the ordinary χ^2 tables is usually adequate. If X^2 has more than 30 degrees of freedom, it tends to become normally distributed, but when the expectations are low, the mean and variance are different from those of the tabular χ^2. Expressions for the exact mean and variance have been given by Haldane (10). Compute the exact mean and variance, and treat X^2 as normally distributed with that mean and variance.*

*In a previous paper (7) I recommended this procedure only when the degrees of freedom in X^2 exceed 60. Some unpublished research suggests that 30 is a better division point.

Further research will presumably change these recommendations, but I do not believe that the recommendations will lead users far astray.

Succeeding sections will deal with some of the common applications of the χ^2 test to goodness of fit problems and to contingency tables. The alternative or supplementary tests to be presented are those that seem most often useful, but they by no means exhaust the possibilities. The important guiding rule is to think about the type of alternative that is likely to hold if the null hypothesis is false, and to select a test that will be sensitive to this kind of alternative.

3. THE GOODNESS OF FIT TEST OF THE POISSON DISTRIBUTION

This is the test already referred to in table 1, except that in practice the parameter m must usually be estimated from the data. We have

$$X^2 = \sum \frac{(f_i - m_i)^2}{m_i}$$

where f_i is the observed and m_i the expected frequency of an observation equal to $i (i = 0, 1, 2 \cdots)$.

If the data do not follow the Poisson distribution, two common alternatives are as follows:

(1) The data follow some other single distribution, such as the negative binomial or one of the "contagious" distributions. Another way of discribing this case is to postulate that the individual observations, say x_i , follow Poisson distributions, but with means that vary from observation to observation so as to follow some fixed frequency distribution. For instance, as has been shown, the negative binomial distribution can be produced by assuming that these means follow a Pearson type III distribution.

(2) The means of the observations x_i follow some systematic pattern. With data gathered over several days, the means might be constant within a day, but vary from day to day, or they might exhibit a slow declining trend.

3.1 *Test of the variance*

In both cases (1) and (2), a comparison between the observed variance of the observations x_i and the variance predicted from Poisson theory will frequently be more sensitive than the goodness of fit test. The variance test is made by calculating

$$X_v^2 = \sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{\bar{x}} ,$$

or if the calculation is made from the frequency distribution of the x_i ,

$$X^2_v = \sum_i \frac{f_i(i - m)^2}{m} ,$$

where $m = \bar{x}$ is the sample mean. The quantity X^2_v is referred to the χ^2 tables with $(N - 1)$ degrees of freedom. The variance test is an old one: it was introduced by Fisher in the first edition of "Statistical Methods for Research Workers" under the heading "Small samples of the Poisson series", because it can be used when the sample is too small to permit use of the goodness of fit test.

The increased power of the variance test over the goodness of fit test was strikingly shown in some sampling experiments conducted by Berkson (20), in a situation in which the data followed a binomial distribution, which has a *smaller* variance than the Poisson. A rough calculation for the example in table 1 gives 0.76 as the probability of rejecting the Poisson hypothesis by the variance test, as compared with 0.56 for the best of the goodness of fit tests. In practice, I have often found the variance test significant when the goodness of fit test was not. Berkson (21) presents some data that illustrate this point.

3.2 *Subdivision of degrees of freedom in the variance test*

If it is suspected that the means of the observations x_i may change in some *systematic* manner, as in case (2) above, more specific tests of significance can be obtained by selecting certain degrees of freedom from X^2_v . The ordinary rules of the analysis of variance are followed in subdividing $\sum(x_i - \bar{x})^2$, and the denominator $\bar{x}$ is used to convert the partial sum of squares approximately to χ^2. A few examples will be given. (The formulas presented are intended to make clear the structure of the χ^2 components, but they are not always the speediest formulas to use in computing the components.)

3.3 *Test for a change in level*

To test for an abrupt change in the mean of the distribution, occurring after the first N_1 observations, we take

$$X^2 = \frac{N_1 N_2}{(N_1 + N_2)} \frac{(\bar{x}_1 - \bar{x}_2)^2}{\bar{x}} \qquad (1 \text{ d.f.})$$

where $N_2 = (N - N_1)$, $\bar{x}_1$, and $\bar{x}_2$ are the sample means in the two parts of the series and $\bar{x}$ is the overall mean. This test may be extended to compare a group of means.

3.4 *Test for a linear trend*

We may anticipate that the means will follow a linear regression on some variate z_i (frequently a time-variable). In this case

$$X^2 = \frac{[\sum (x_i - \bar{x})(z_i - \bar{z})]^2}{\bar{x} \sum (z_i - \bar{z})^2} \qquad (1 \text{ d.f.})$$

3.5 *Detecting the point at which a change in level occurs*

This problem has been illustrated by Lancaster (11) in an experiment in which increasing concentrations of disinfectant are poured on a series of suspensions of a bacterial culture, each suspension having a constant amount of bacteria. The observations x_i represent numbers of colonies per plate. The problem is to find the value of j for which the disinfectant is strong enough to begin reducing the number of colonies. To this end we compare each observation with the mean of all previous observations, looking for the first value of j at which X^2 becomes large owing to a drop in the number of colonies. We thus obtain a set of independent single degrees of freedom:

$$X_1^2 = \frac{(x_1 - x_2)^2}{2\bar{x}} ; \qquad X_2^2 = \frac{(x_1 + x_2 - 2x_3)^2}{6\bar{x}} ;$$

and in general

$$X_r^2 = \frac{(x_1 + x_2 + \cdots + x_r - rx_{r+1})^2}{r(r+1)\bar{x}}$$

Note. The above set of X^2 values will add up to the total variance X_t^2, but as Lancaster has pointed out, there are other natural ways of subdividing X_t^2 in which the separate X^2 do not add up to X_t^2. When comparing x_1 with x_2, we might decide to disregard the remainder of the observations and compute X_1^2 as

$$X_1'^2 = \frac{(x_1 - x_2)^2}{2\bar{x}_{12}}$$

where $\bar{x}_{12}$ is the mean of x_1 and x_2. A set of successive X^2 values computed in this way will not add up to X_t^2, because the denominator changes from term to term, whereas X_t^2 carries the denominator $\bar{x}$.

The practice of computing X^2 components by using only those parts of the data that are immediately involved has something in its favor (despite the non-additive feature), at least if the total X_t^2 has already been shown to be significant. For in that event we have already concluded that the data as a whole do not follow a single Poisson distribution, and the overall mean $\bar{x}$ is of dubious validity as an estimate

of the Poisson variance for a part of the data. On the other hand, if the total X^2_7 is not significant, but we suspect that some component is, the additive partition is convenient and should be satisfactory in a preliminary examination.

3.6 *Single degrees of freedom in the goodness of fit test*

In the goodness of fit comparison of the observed frequency f_i with the expected frequency m_i of i occurrences, we can test any linear function of the deviations

$$L = \sum g_i(f_i - m_i)$$

where the g_i are numbers chosen in advance.

In the case in which the mean of the Poisson, and hence the m_i , are given in advance, the variance of L is

$$V(L) = \sum g_i^2 m_i - \frac{(\sum g_i m_i)^2}{N} \tag{1}$$

In the more common situation in which the Poisson mean m is estimated from the data,

$$V(L) = \sum g_i^2 m_i - \frac{(\sum g_i m_i)^2}{N} - \frac{[\sum g_i m_i(i - m)]^2}{Nm} \tag{2}$$

where the sums are over the values 0, 1, 2, $\cdots$ of i. In either case

$$X_1^2 = \frac{L^2}{V(L)} \tag{3}$$

is approximately distributed as χ^2 with 1 d.f. I plan to publish justification for formula (2), which appears to be new. By appropriate choice of the g_i , a test specific for a given pattern of deviations is obtained.

In particular, to test any single deviation $(f_i - m_i)$ when m is estimated from the data, we take

$$L = (f_i - m_i) : V(L) = m - \frac{m_i^2}{N}\left\{1 + \frac{(i - m)^2}{m}\right\} \tag{4}$$

As an illustration, the data in table 2 are for a sample which gave a satisfactory fit to a Poisson distribution. However, in copying down the frequencies before fitting the Poisson, the frequency of 3 occurrences was erroneously written as 52 instead of 32.

TABLE 2.
GOODNESS OF FIT TEST FOR A SAMPLE WITH A GROSS ERROR.

i	f_i	m_i	Contribution to X^2
0	52	47.65	0.40
1	67	77.04	1.31
2	58	62.28	0.29
3	52	33.56	10.13
4	7	13.56	3.17
5	3	4.39	0.44
6+	1	1.52	0.18
Total	240	240.00	15.92

The value of m is 1.6167. The total X^2, 15.92 with 5 d.f., is significant at the 1 per cent level. The large contribution to X^2 from $i = 3$ excites notice. In order to test this deviation, we take

$$L = f_3 - m_3 = 52 - 33.56 = 18.44$$

$$V(L) = 33.56 - \frac{(33.56)^2}{240}\left\{1 + \frac{(3 - 1.6167)^2}{1.6167}\right\} = 23.31$$

$$X_1^2 = \frac{(18.44)^2}{23.31} = 14.59$$

This comparison accounts for the major part of the total X^2. It must be pointed out, however, that the X_1^2 test applies only to a deviation picked out before seeing the data. Thus the test can be applied validly, for $i = 0$, say, if we suspect beforehand that the data follow the Poisson distribution for $i \geq 1$, but that the frequency for $i = 0$ may be anomalous. If the test is applied, as here, to a deviation selected because it looks abnormally large, the significance P obtained is too low. I do not have an expression for the correct significance probability when we select the largest deviation. It appears intuitively that the correct probability lies between P and kP, where k is the number of classes in the goodness of fit test. Since P is about 0.00013 and k is 7, the upper limit is .00091, which is still highly significant statistically.

4. THE GOODNESS OF FIT TEST OF THE BINOMIAL DISTRIBUTION

For the binomial distribution, there is a series of tests analogous to those given for the Poisson distribution. A typical observation con-

sists of the number of successes x_i out of n independent trials. We have a sample of N such observations. The ordinary goodness of fit test is made by recording the frequency f_i with which i successes occur in the sample, fitting a binomial to these frequencies, and calculating X^2 as

$$X^2 = \sum \frac{(f_i - m_i)^2}{m_i},$$

where m_i is the corresponding expected frequency.

As in the Poisson case, departures from the binomial frequently occur either because

(1) the data follow a different frequency distribution, usually with a larger variance (or the probabilities of success p_i show some kind of random variation from observation to observation).

(2) the probabilities p_i are affected by a systematic source of variation.

In both cases, a comparison of the observed and expected variances is likely to be more sensitive than the goodness of fit test.

4.1 *Test of the variance*

The test criteria, all distributed approximately as χ^2, are as follows.

n constant, p given in advance

$$X_v^2 = \frac{\sum_j (x_j - np)^2}{npq} = \frac{\sum_i f_i(i - np)^2}{npq}, \qquad N \text{ d.f.}$$

n constant, p estimated

$$X_v^2 = \frac{\sum_j (x_j - \bar{x})^2}{n\hat{p}\hat{q}} = \frac{\sum_i f_i(i - n\hat{p})^2}{n\hat{p}\hat{q}}, \qquad (N - 1) \text{ d.f.}$$

where $n\hat{p} = \bar{x}$, and $\hat{q} = 1 - \hat{p}$.

n varying, p given in advance

In this case we cannot make a simple goodness of fit test (unless the sample is large enough to be divided into batches, each with n constant, so that the test can be made separately for each batch). If $p_j = x_j/n_j$ is the observed proportion of successes in the jth member of the sample,

$$X_v^2 = \frac{\sum n_j(p_j - p)^2}{pq} \qquad N \text{ d.f.}$$

n varying, p estimated

$$X_v^2 = \frac{\sum n_i(p_i - \hat{p})^2}{\hat{p}\hat{q}} = \frac{\sum \frac{x_i^2}{n_i} - \frac{(\sum x_i)^2}{(\sum n_i)}}{\hat{p}\hat{q}}, \qquad (N - 1) \text{ d.f.}$$

where $\hat{p} = (\sum x_i)/(\sum n_i)$ is the estimate of p from the total sample.

There is another way of deriving the variance test. Arrange the data in a $2 \times N$ contingency table, as follows.

Successes	x_1	x_2		x_N
Failures	$n_1 - x_1$	$n_2 - x_2$		$n_N - x_N$
Total	n_1	n_2		n_N

Then the X^2 that is used to test for association in this $2 \times N$ table may be shown to be identical with X_v^2.

If N exceeds 30 and the expectations are small, it was pointed out in section 2 that X^2 in contingency tables tends to a normal distribution with a mean and variance somewhat different from those of χ^2. Use of Haldane's correct expressions for the mean and variance of X^2 was recommended in this case. The same procedure is recommended in the variance test if N exceeds 30 and the average n_i is less than 10. Haldane's expressions are rather complicated when the n_i vary. In the fairly common situation in which n is constant and p is estimated from the data, the following results suffice in almost all cases.

$$E(X_v^2) \doteq (N - 1)\left(1 + \frac{1}{Nn}\right) \tag{5}$$

$$V(X_v^2) \doteq 2(N - 1)\left(\frac{n - 1}{n}\right)\left(1 - \frac{1 - 7\hat{p}\hat{q}}{Nn\hat{p}\hat{q}}\right) \tag{6}$$

The important correction term is that in $(n - 1)/n$ in the variance: the terms in $1/Nn$ are usually small. These results will be used in the numerical example which follows.

The data in table 3, taken from a previous paper, (12), illustrate the application of the goodness of fit test and the variance test to the same observations. The original data, due to Dr. J. G. Bald, consisted of 1440 tomato plants in a field having 24 rows with 60 plants in a row. For each plant it was recorded whether the plant was healthy or attached by spotted wilt as of a given date. As one method of examining whether the distribution of diseased plants was random over the field, the plants were divided into 160 groups of 9, each group consisting of 3

plants × 3 rows. Thus $N = 160$, $n = 9$. The choice of $n = 9$ was of course arbitrary, and I do not know what would have been the best choice. The obvious alternative to a binomial distribution is that the values of p_i vary from one group of 9 to another, indicating a patchiness in distribution.

TABLE 3.
NUMBERS OF DISEASED PLANTS IN GROUPS OF 9 PLANTS.

i	f_i	m_i	Contr. to X^2
0	36	26.45	3.45
1	48	52.70	0.42
2	38	46.67	1.61
3	23	24.11	0.05
4	10	8.00	0.50
5	3		
6	1	2.05	4.25
7	1		
8	0		
	N = 160	159.98	10.28

Allowing a minimum expectation of 1, we must pool the last 4 classes in table 3. The value of X^2 is 10.28, with 4 d.f., since p is estimated; the significance P is 0.036. (If we pooled the last 5 classes in order to have a minimum expectation of 5, we would obtain a significance P of 0.046.)

For the variance test, we compute X_s^2 from the observed frequency distribution in table 3. We have

$$\sum if_i = 261 = Nn\hat{p}, \qquad \text{so that } \hat{p} = 0.18125.$$

$$X_s^2 = \frac{\sum i^2 f_i - (\sum if_i)^2/N}{n\hat{p}\hat{q}} = \frac{727 - (261)^2/160}{9(0.18125)(0.81875)}$$

$$= 225.55, \text{ with } 159 \text{ d.f.}$$

Since $N = 160$ and $n = 9$, we use the normal approximation to the distribution of X_s^2, based on the correct mean and variance. In expressions (5) and (6), the terms in $1/Nn$ are negligible ($Nn = 1440$). Hence we take

$$E(X_s^2) = 159 : V(X_s^2) = 2(159)(8)/9 = 282.66$$

The approximate normal deviate is

$$\frac{225.55 - 159}{\sqrt{282.66}} = 3.96$$

This has a significance P less than 0.0001, much lower than that obtained from the goodness of fit tests.

Subdivision of the sum of squares for $X^2_{\cdot}$, which may be useful in testing for systematic variation of the p_i , will be discussed in section 6, which deals with $2 \times N$ contingency tables.

4.2 *Single degrees of freedom in the goodness of fit test*

Let

$$L = \sum g_i(f_i - m_i)$$

be a specified linear function of the deviations of observed from expected frequencies. If p is given, formula (1) in section 3.6 holds for the variance of L. If p is estimated from the data,

$$V(L) = \sum g_i^2 m_i - \frac{(\sum g_i m_i)^2}{N} - \frac{[\sum g_i m_i(i - n\hat{p})]^2}{Nn\hat{p}\hat{q}} \qquad (7)$$

For a single deviation, $(f_i - m_i)$, selected in advance,

$$V(L) = m_i - \frac{m_i^2}{N}\left\{1 + \frac{(i - n\hat{p})^2}{n\hat{p}}\right\} \qquad (8)$$

Then $L^2/V(L)$ is approximately distributed as χ^2 with 1 d.f.

5. THE GOODNESS OF FIT TEST OF THE NORMAL DISTRIBUTION

When the normal distribution is fitted to a body of data, both the mean and the variance are estimated from the sample: consequently, no variance test is possible. However, the variance test is just an application of the general procedure in which we compare the lowest moments (or cumulants) in which the sample can differ from the theoretical distribution that is being fitted. In this sense, the analogue of the variance test is the test for skewness (as given e.g. in Fisher's "Statistical Methods for Research Workers," §14), which will often detect a departure from normality that escapes the goodness of fit test. The test for kurtosis is also useful in this connection.

As with the Poisson and binomial distributions, a specified linear function L of the deviations in the goodness of fit test can be scrutinized. The formula for the variance of L is somewhat more complicated,

because the observed frequencies are subject to 3 constraints. If

$$L = \sum g_i(f_i - m_i)$$

then

$$V(L) = \sum g_i^2 m_i - \frac{(\sum g_i m_i)^2}{N} - \frac{(\sum g_i d_i m_i)^2}{Ns^2} - \frac{[\sum g_i m_i(d_i^2 - s^2)]^2}{2Ns^4}$$

where

d_i = (midpoint of ith class) − (sample mean)

s^2 = sample estimate of variance

In order to apply this test, construct 3 additional columns containing the quantities $g_i m_i$, d_i and $(d_i^2 - s^2)$, respectively. $V(L)$ is then easily computed.

For testing a *single* deviation, $V(L)$ simplifies to

$$V(L) = m_i - \frac{3m_i^2}{2N} - \frac{m_i^2 d_i^4}{2Ns^4}$$

The series of supplementary tests described above for the Poisson, binomial and normal distributions can be extended to other distributions. In particular, variance and skewness tests for the negative binomial distribution have been developed by Anscombe (18) and further illustrated by Bliss (19).

6. SUBDIVISION OF DEGREES OF FREEDOM IN THE 2 x N CONTINGENCY TABLE

This section describes some useful ways in which the total X^2 for a $2 \times N$ contingency table may be subdivided. The notation, which continues that already used for the binomial distribution, is as follows.

	Number of			Proportion of Successes
	Successes	Failures	Total	
	x_1	$n_1 - x_1$	n_1	$p_1 = x_1/n_1$
	x_2	$n_2 - x_2$	n_2	$p_2 = x_2/n_2$
	- - - -	- - - - - - -	- - - - - - -	- - - - - - - -
	x_N	$n_N - x_N$	n_N	$p_N = x_N/n_N$
Totals	T_x	$T - T_x$	T	$\bar{p} = T_x/T$

For many purposes, a formula due to Brandt and Snedecor is useful in interpreting the total X^2. The formula is:

$$X^2 = \frac{\sum_{i=1}^{N} n_i(p_i - \hat{p})^2}{\hat{p}\hat{q}} \tag{9}$$

Thus the total X^2 is seen to be a weighted sum of squares of the deviations of the individual proportions of success p_i from their mean, with weights $n_i/\hat{p}\hat{q}$. Consequently, if we subdivide this weighted sum of squares into a set of independent components by the rules of the analysis of variance, we obtain a corresponding subdivision of X^2 into independent components.

6.1 *Test for a change in the level of p*

In order to test whether the value of p is different in the first N_1 rows from that in the subsequent N_2 rows ($N = N_1 + N_2$), we may subdivide X^2 into the following 3 components.

	d.f.
Difference between p's in first N_1 and last N_2 rows	1
Variation among p_i within the first N_1 rows	$(N_1 - 1)$
Variation among p_i within the last N_2 rows	$(N_2 - 1)$

For the following example I am indebted to Dr. Douglas P. Murphy. A group of women known to have cancer of the uterus and a corresponding 'control' group (primarily from a dental clinic and several women's clubs) were selected. From each of a defined set of relatives of the selected person (the proband), data were secured about the presence of cancer. A higher proportion of cancer cases among relatives of the cancer proband would indicate some kind of familial aggregation of the disease, perhaps of genetic origin (13).

In one table, data were presented separately for those relatives who were of the same generation as the proband (e.g. sister) and for those relatives who were one generation earlier (e.g. mother). Some breakdown of this kind is advisable, because cancer attacks mainly in middle or old age, and the 'cancer' and 'control' groups might be found to differ in the proportions of young relatives which they contained. The data, which are a small part of a much more intensive investigation, appear in table 4.

There are several ways of computing the total X^2 in a $2 \times N$ table. One form of the Brandt-Snedecor formula is

$$X^2 = \frac{\sum x_i p_i - \hat{p}T_x}{\hat{p}\hat{q}} \tag{10}$$

TABLE 4.
CANCER AMONG RELATIVES OF 'CANCER' AND 'CONTROL' PROBANDS

	Proband	No. of relatives			Proportion with cancer
		With x_i	Without	Total n_i	p_i
Earlier Generation	Cancer	86	814	900	0.095556
	Control	117	1038	1155	0.101299
Same Generation	Cancer	49	1475	1524	0.032152
	Control	61	1580	1641	0.037172
	Total	313	4907	5220	$\hat{p} = 0.059962$ $\hat{q} = 0.940038$ $\hat{p}\hat{q} = 0.056367$

This expression is useful when there is some question of a systematic variation in the p_i , because we will want to compute the p_i in order to have a look at them. When the p_i have been computed, the numerator of X^2 can be obtained from formula (10) directly on the computing machine, without any intermediate writing down. The only disadvantage is that a substantial number of decimal places must be retained in the p_i .

If this method is to be used, first compute $\hat{p}$, $\hat{q}$ and the product $\hat{p}\hat{q}$ (bottom right of table 4). Since $\hat{p}\hat{q}$ is about 1/20, the numerator of X^2 must be correct to at least 3 decimal places if we want X^2 correct to 2 decimal places. Further, since T_x is 313, we should have 6 decimal places in the p_i . (The symbol x_i should be assigned to the column with the *smaller* numbers: this makes the computations lighter and necessitates fewer decimals in the p_i).

From inspection of the p_i in table 4, a large difference in cancer rates between the two generations is evident. Within each generation, the differences in rates between the cancer and control groups appear tiny. To illustrate the methods, the total X^2 will be partitioned into the 3 relevant components. All that is necessary is to subdivide the sum of squares in the numerator, and then divide each component by $\hat{p}\hat{q}$. For the numerator of the total X^2, we have from formula (10),

$$\text{Total } S.S. = (86)(0.095556) + \cdots + (61)(0.037172) - (313)(0.059962)$$
$$= 5.1447$$

For the comparison of the two generations, we form the auxiliary 2×2 table.

	With x_i	Without	Total n_i	Proportion p_i
Earlier generation	203	1852	2055	0.098783
Same generation	110	3055	3165	0.034755
Total	313	4907	5220	0.055962

The same formula can be used for this table.

$$\text{Generations } SS = (203)(0.098783) + (110)(0.034755) - (313)(0.055962) = 5.1079.$$

For the comparisons between cancer and control groups within generations, we have

$$\text{First generation } SS = (86)(0.095556) + (117)(0.101299) - (203)(0.098783) = 0.0169.$$

The second generation SS, obtained similarly, gives 0.0199. In table 5 the results are summarized and converted to X^2 values on division by $\hat{p}\hat{q}$.

TABLE 5.
SUBDIVISION OF X^2 INTO COMPONENTS (2 x 4 TABLE)

Component	d.f.	S.S.	X^2	From 2 x 2 tables
First vs. later generation	1	5.1079	90.62	90.62
Cancer vs. Control: first gen.	1	0.0169	0.30	0.19
Cancer vs. Control: later gen.	1	0.0199	0.35	0.59
Total	3	5.1447	91.27	91.40

There is no indication of any difference in cancer rates between the cancer and control groups within either generation. In order to complete this analysis, we should make a combined test of Cancer vs. Control from the two generations. Methods for making tests of this kind are discussed in section 8.

As mentioned previously in connection with the corresponding subdivision for the Poisson distribution, separation of X^2 into *additive*

components is convenient for a preliminary examination of the data. But if differences in p are found between groups of rows, the X^2 values for comparisons made *within* groups need to be recomputed. This is clear in the present example. The additive method requires the assumption that the estimated variance of any p_i is $\hat{p}\hat{q}/n_i$. However, the huge X^2 value of 90.62 between generations shows that the combined $\hat{p}$ cannot be regarded as a valid estimate of the proportion of cancer cases within either of the individual generations.

The procedure is to recompute the two 'within-generation' X^2, each from its own 2×2 table. These values, which no longer add to the original total X^2, are given in the right hand column of table 5. The interpretation of the data is not altered. The X^2 values computed from individual parts of the table seldom differ greatly from the additive X^2, but they can do so in certain circumstances and are worth looking at, as a precaution, in analyses of this type.

In this example the rows were divided into 2 groups. The same methods may be applied to test the variation in p among any number of groups, and also within each group. To obtain an additive separation, we subdivide the numerator of X^2 as indicated in the example. Alternatively, for a non-additive separation, we can form an auxiliary table in which each row is a group total and obtain the X^2 for this table, and a further X^2 for each group considered by itself.

6.2 *Test for a linear regression of p.*

In some contingency tables we may expect that the p_i will bear a linear relation to a variate z_i that is defined for each of the N rows of the table. In others, where the rows fall into a natural order, it is not unreasonable to assign scores z_i to the rows, in an attempt to convert the ordering into a continuous scale, with which the p_i may show a linear relation. Since p_i is assigned a weight $n_i/\hat{p}\hat{q}$, the regression coefficient b of p_i on z_i is obtained by the standard formula for weighted regressions:

$$b = \frac{\sum n_i(p_i - \hat{p})(z_i - \bar{z}_w)}{\sum n_i(z_i - \bar{z}_w)^2} \tag{11}$$

where $\bar{z}_w$ is the weighted mean of the z_i .

For computing purposes, the numerator and denominator of b are conveniently expressed as follows (note that $n_i p_i = x_i$):

$$\text{Num.} = \sum x_i z_i - \frac{T_z(\sum n_i z_i)}{T} \tag{12}$$

$$\text{Den.} = \sum n_i z_i^2 - \frac{(\sum n_i z_i)^2}{T} \tag{13}$$

The X^2 for regression, with 1 d.f., is

$$X^2 = \frac{(\text{Num.})^2}{\hat{p}\hat{q}(\text{Den.})} \tag{14}$$

As an illustration, the data in table 6, for which I am indebted to the Leonard Wood Memorial, (American Leprosy Foundation) are taken from an experiment on the use of drugs (sulfones and streptomycin) in the treatment of leprosy. The rows denote the *change* in the overall clinical condition of the patient during 48 weeks of treatment: the columns indicate the degree of infiltration (a measure of a certain type of skin damage) present at the beginning of the experiment. The question of interest is whether patients with much initial infiltration progressed differently from those with little infiltration. Patients did not all receive the same drugs, but since no difference in the effects of drugs could be detected, it was thought that the data for different drugs could be combined for this analysis.

TABLE 6.
196 PATIENTS CLASSIFIED ACCORDING TO CHANGE IN CONDITION AND DEGREE OF INFILTRATION

Clinical change		Score z_i	Degree of infiltration		Total n_i	$p_i = x_i/n_i$ (in %)	$n_i z_i$
			0 − 7	8 − 15 x_i			
Improvement	Marked	3	11	7	18	39	54
	Moderate	2	27	15	42	36	84
	Slight	1	42	16	58	28	58
	Stationary	0	53	13	66	20	0
	Worse	−1	11	1	12	8	−12
Total			144	52 T_x	196 T	0.26531 $\hat{p}$	184

The total X^2 is 6.88, with 4 d.f. (P about 0.16). However, the p_i (the proportions of patients with severe infiltration) decline steadily from 39% in the "markedly improved" class to 8% in the "worse" class. This suggests that a regression of the p_i on the clinical change might furnish a more sensitive test.

The data are typical of many tables in that the rows (clinical changes) are ordered. In order to compute a regression, this ordering must be replaced by a numerical scale. I have supposed that scores 3, 2, 1,

0, −1, as shown in table 6, have been assigned to the five classes of clinical change. Such scores are to some extent subjective and arbitrary, and some scientists may feel that the assignment of scores is slightly unscrupulous, or at least they are uncomfortable about it. Actually, any set of scores gives a *valid* test, provided that they are constructed without consulting the results of the experiment. If the set of scores is poor, in that it badly distorts a numerical scale that really does underlie the ordered classification, the test will not be sensitive. The scores should therefore embody the best insight available about the way in which the classification was constructed and used. In the present example, I considered an alternative set 4, 2, 1, 0, −2, on the grounds that the doctor seemed to deliberate very carefully before assigning a patient to the "markedly improved" or to the "worse" class, but I decided that the presumption in favor of this scale was not strong enough.

To compute the regression X^2, the only supplementary column needed is that of the products $n_i z_i$, shown on the right of table 6. From equations (12) and (13) for b,

$$\text{Num.} = \sum x_i z_i - \frac{T_x(\sum n_i z_i)}{T}$$

$$= (7)(3) + (15)(2) + \cdots + (1)(-1) - \frac{(52)(184)}{196}$$

$$= 17.1837 \qquad (15)$$

$$\text{Den.} = \sum n_i z_i^2 - \frac{(\sum n_i z_i)^2}{T}$$

$$= (54)(3) + (84)(2) + \cdots + (-12)(-1) - \frac{(184)^2}{196}$$

$$= 227.2653 \qquad (16)$$

$$X^2 \text{ for regression} = \frac{(17.1837)^2}{(227.2653)\hat{p}\hat{q}} = 6.666 \qquad (1 \text{ d.f.})$$

This is significant at the 1% level. The total X^2 has now been subdivided as follows.

	d.f.	X^2
Regression of p_i on z_i	1	6.67
Deviations from regression	3	0.21
Total	4	6.88

6.3 *Comparison of mean scores.*

There is another way of looking at the relation between degree of infiltration and clinical progress. We might ask whether the degree of improvement is, on the average, better for patients with severe infiltration than for patients with less infiltration. In many applications, including the present one, this is a more natural way of posing the scientific question than by asking whether the proportion of patients with severe infiltration changes from class to class.

Yates (14) has shown how to compute X^2, with 1 d.f., for comparing the mean scores, and has proved, in a more general case, that this X^2 is identical with the regression X^2 which we have just considered. In the 'mean score' approach, it is best to think of the two groups of patients as representing two independent samples. Each patient has been assigned a variate or score z_i , and the x_i now represent the frequencies with which these scores occur in one of the samples. The mean score for a sample is

$$\bar{z} = \frac{\sum x_i z_i}{\sum x_i} = \frac{\sum x_i z_i}{T_x}$$

In table 6, this has the value 0.8194 for patients with 0–7 degrees of infiltration and 1.2692 for patients with 8–15 degrees. The formula for the variance of a mean score is, from Yates (14),

$$V(\bar{z}) = \frac{1}{T_x T}\left[\sum z_i^2 n_i - \frac{(\sum z_i n_i)^2}{T}\right] \qquad (17)$$

Note that, except for the term in T_x , this formula is exactly the same for each of the two means.

The value of the term in square brackets has already been obtained in equation (16) as 227.2653. Hence, from equation (17) and the data in table 6,

$$V(\bar{z}_1) = \frac{227.2653}{(144)(196)} : V(\bar{z}_2) = \frac{227.2653}{(52)(196)}$$

$$V(\bar{z}_1 - \bar{z}_2) = \frac{227.2653}{196}\left\{\frac{1}{144} + \frac{1}{52}\right\} = 0.03035$$

The X^2 value for the difference in mean scores is

$$X^2 = \frac{(\bar{z}_1 - \bar{z}_2)^2}{V(\bar{z}_1 - \bar{z}_2)} = \frac{(1.2692 - 0.8194)^2}{0.03035} = 6.666$$

in agreement with the regression X^2.

6.4 *Step by step comparisons of the* p_i .

Occasionally it is useful to compare p_1 with p_2 , the weighted mean of p_1 and p_2 with p_3 , and so on. To obtain an additive subdivision of X^2, we partition the sum of squares in the numerator of X^2 as follows:

$$\frac{n_2\{n_1p_1 - n_1p_2\}^2}{n_1(n_1 + n_2)},$$

$$\frac{n_3\{n_1p_1 + n_2p_2 - (n_1 + n_2)p_3\}^2}{(n_1 + n_2)(n_1 + n_2 + n_3)},$$

and for the general term,

$$\frac{n_{r+1}\{n_1p_1 + \cdots + n_rp_r - (n_1 + \cdots + n_r)p_{r+1}\}^2}{(n_1 + \cdots + n_r)(n_1 + \cdots + n_{r+1})}$$

Each term is then divided by $\hat{p}\hat{q}$ to convert it to X^2 with 1 d.f.

For the corresponding non-additive subdivision, we calculate X^2, by the standard methods, for each of the following set of 2×2 tables.

x_1	y_1	n_1
x_2	y_2	n_2
$x_1 + x_2$	$y_1 + y_2$	

$x_1 + x_2$	$y_1 + y_2$	$n_1 + n_2$
x_3	y_3	n_3
$x_1 + x_2 + x_3$	$y_1 + y_2 + y_3$	

and so on, where we have written y_i for $(n_i - x_i)$.

7. THE GENERAL TWO-WAY CONTINGENCY TABLE

For considerations of space, this case will be discussed very briefly. The following notation will be used for the frequencies in the $r \times c$ cells of a table with c columns and r rows.

					Total
	x_{11}	x_{21}		x_{c1}	n_1
	x_{12}	x_{22}		x_{c2}	n_2
	x_{1r}	x_{2r}		x_{cr}	n_r
Total	T_1	T_2		T_c	T

One fairly common situation is that the rows can be divided into g groups. We expect that within any group there is no association between rows and columns, but there is reason to examine for association

among the group totals. To put it another way, the ratios of the p_{ij} from column to column may be constant within a group but vary from group to group. For this situation, a non-additive separation of the total X^2 is obtained by: (1), dividing the contingency table into the appropriate g contingency tables, and calculating the X^2 in the ordinary way for each separate table; (2), forming a new $g \times c$ table in which the entries are column totals within each group, and calculating the ordinary X^2 for this table.

Lancaster (11) and Irwin (15) have shown how to make a complete (non-additive) separation into single d.f. by subdividing the table into a set of 2×2 contingency tables, $(r - 1)(c - 1)$ in number, and computing X^2 in the ordinary way for each 2×2 table. The method of forming the 2×2 tables is easily grasped. The same authors have given general methods for the production of additive subdivisions of the total X^2. In the case of the partition into single d.f. just mentioned, Kimball (16) gives an expeditious method of computing the individual X^2 values. As in the $2 \times N$ case, the basic difference between additive and non-additive partitions is that in the former the estimated variances used to convert the sums of squares into approximate χ^2 values are obtained from the margins of the whole table, while in the non-additive partition they are obtained separately from the margins of each part.

7.1 *Score methods.*

If the rows, say, represent an ordered classification, and scores z_i can be assigned to them, we can compute mean scores

$$\bar{z}_i = \frac{\sum_i z_i x_{ij}}{T_j} = \frac{U_j}{T_j} \qquad (j = 1, 2, \cdots, c)$$

for each column, where U_j may be called the total score for the jth column. From equation (17), section 6.3, noting that T_x is now replaced by T_j in our notation, we have

$$V(\bar{z}_j) = \frac{1}{T_j T}\left[\sum z_i^2 n_i - \frac{(\sum z_i n_i)^2}{T}\right] = \frac{D}{T_j}, \qquad (17)'$$

where the factor D is the same for all columns.

These results provide a test of the hypothesis that there is no difference in mean score from column to column to column. We have

$$X^2 = \frac{\sum T_j(\bar{z}_j - \bar{z})^2}{D} = \frac{\sum \frac{U_j^2}{T_j} - \frac{(\sum U_j)^2}{T}}{D}. \qquad (18)$$

The second form is better for computing. The d.f. are $(c - 1)$. These d.f. may in turn be subdivided for more specific comparisons among the columns. In particular, if scores have also been assigned to each column, we may test the regressions of the column means on these scores. General methods have been given by Yates (14).

7.2 *Use of the analysis of variance.*

For the reader who is familiar with the analysis of variance, but less so with χ^2 tests in contingency tables, it is worth noting that the analyses based on scores may be performed quite satisfactorily by ordinary analysis of variance methods. The approach can be illustrated from the comparison of the mean scores of the two groups of patients which was made from the data in table 6.

Let us regard the data in table 6 as representing two independent samples of data from frequency distributions in which the variates take only the values 3, 2, 1, 0, -1. The orthodox analysis of variance, using a pooled estimate of error, is given in table 7.

TABLE 7.
COMPARISON OF MEAN SCORES BY ANALYSIS OF VARIANCE

Source of variation	d.f.	S.S.	M.S.
Between sample means	1	7.7289	7.7289
Within samples	194	219.5364	1.1316

The F-value is 6.830 with 1 and 194 d.f., as compared with the X^2 value of 6.666, which we may regard as an F-value of 6.666 with 1 and ∞ d.f. Clearly, the significance probabilities differ by a negligible amount.

For an $r \times c$ contingency table, the general relation between the F and X^2 values for comparing the mean scores for the c columns is

$$F = \frac{X^2}{df} \cdot \frac{(T - c)}{(T - X^2)} \tag{19}$$

where df is the number of degrees of freedom in X^2.

This relation may be verified numerically in the present example. Since $X^2 = 6.666$, $df = 1$, $T = 196$, $c = 2$, we obtain 6.830 for the right side of equation (19), in agreement with the actual F-value. In most tables, T will be much larger than either c or X^2, so that F and X^2 will be practically equal. Further, since F usually has a substantial

number of d.f. in the denominator, the significance probabilities obtained from F and X^2 will usually agree closely.

The F-test has one aesthetic objection. As Yates showed, the X^2 test gives *exactly* the same result, as it should, for the regression of the p_i on the row scores as for the comparison of the mean scores in the two columns. The F-test can also be performed both ways. We can assign a score 0 to one column and a score 1 to the other, and make an F-test of the regression of the *row* means on the row scores. This F-value, however, is not identical with the F-value obtained in table 7. However, the significance levels differ only by trivial amounts except when T is small.

8. THE COMBINATION OF 2 x 2 CONTINGENCY TABLES

Suppose that we are comparing the frequencies of some occurrence in two independent samples, and that the whole procedure is repeated a number of times under somewhat differing environmental conditions. The data then consist of a series of 2×2 tables, and the problem is to make a combined test of significance of the difference in occurrence rates in the two samples. The data obtained in comparing the effectiveness of two agents in dosage-mortality experiments are a typical example, in which the repetitions of the experiment are made under a series of different dosage levels. My concern here, however, is with cases where there is no variate corresponding to dosage level, and no well-established theory of how to combine the data.

One method that is sometimes used is to combine all the data into a single 2×2 table, for which X^2 is computed in the usual way. This procedure is legitimate only if the probability p of an occurrence (on the null hypothesis) can be assumed to be the same in all the individual 2×2 tables. Consequently, if p obviously varies from table to table, or we suspect that it may vary, this procedure should not be used.

Another favorite technique is to compute the usual X^2 separately for each table, and add them, using the fact that the sum of g values of χ^2, each with 1 d.f., is distributed as χ^2 with g d.f. This is a poor method. It takes no account of the signs of the differences $(p_1 - p_2)$ in the two samples, and consequently lacks power in detecting a difference that shows up consistently in the same direction in all or most of the individual tables.

An alternative is to compute the X values, and add them, taking account of the signs of the differences. Since X is approximately normally distributed with mean 0 and unit S.D., the sum of g independent X values is approximately normally distributed with mean 0

and S.D. $\sqrt{g}$. Hence the test criterion,

$$\frac{\sum X}{\sqrt{g}}$$

is referred to the standard normal tables.

This method has much to commend it if the total N's of the individual tables do not differ greatly (say by more than a ratio of 2 to 1) and if the p's are all in the range 20-80%. For the following illustrative data I am indebted to Dr. Martha Rogers. The comparison is between mothers of children in the Baltimore schools who had been referred by their teachers as presenting behavior problems, and mothers of a comparable group of control children who had not been so referred. For each mother, it was recorded whether she had suffered any infant losses (e.g. stillbirths) previous to the birth of the child in the study. The comparison is part of a study of possible associations between behavior problems in children and complications of pregnancy of the mother. Since these loss rates increase with later birth orders, and since the samples might not be comparable in birth orders, the data were examined separately, as a precaution, for 3 birth-order classes (see table 8). The two groups of children are referred to as 'Problems' and 'Controls'.

TABLE 8.
DATA ON NUMBER OF MOTHERS WITH PREVIOUS INFANT LOSSES

Birth Order		No. of mothers with		Total	% Loss
		Losses	None		
2	Problems	20	82	102	19.6
	Controls	10	54	64	15.6
		30	136	166 = N_1	18.1
3–4	Problems	26	41	67	38.8
	Controls	16	30	46	34.8
		42	71	113 = N_2	37.2
5+	Problems	27	22	49	55.1
	Controls	14	23	37	37.8
		41	45	86 = N_3	47.7

Note that the loss rate is higher in the 'Problems' sample in all 3 tables. Since the N's in the separate tables lie within a 2:1 ratio, and the p's are between 18% and 48%, addition of the X values is indicated. The individual X values are, respectively, 0.650, 0.436, 1.587, all being given the same sign since the difference is in the same direction. For this test, the X values are computed without the correction for continuity. Hence the test criterion is the approximate normal deviate

$$\frac{0.650 + 0.436 + 1.587}{\sqrt{3}} = 1.54$$

The P value is just above 0.10. Addition of the X^2 values gives 3.131, with 3 d.f., corresponding to a P of about 0.38.

If the N's and p's do not satisfy the conditions mentioned, addition of the X's tends to lose power. Tables that have very small N's cannot be expected to be of much use in detecting a difference, yet they receive the same weight as tables with large N's. Where differences in the N's are extreme, we need some method of weighting the results from the individual tables. Further, if the p's vary from say 0 to 50%, the difference that we are trying to detect, if present, is unlikely to be constant at all levels of p. A large amount of experience suggests that the difference is more likely to be constant on the probit or logit scale. As a further complication, the term pq in the variance of a difference will change from one 2×2 table to another.

Perhaps the best method for a combined analysis is to transform the data to a probit or logit scale. Examples of this type of analysis are given by Winsor (17) and Dyke and Patterson (22): it is recommended if the data are extensive enough to warrant a searching examination. As an alternative, the following test of significance in the original scale will, I believe, be satisfactory under a wide range of variations in the N's and p's from table to table.

For the ith 2×2 table, let

n_{i1} , n_{i2} = sample sizes
p_{i1} , p_{i2} = observed proportions in the two samples
$\hat{p}_i$ = combined proportion from the margins
d_i = $p_{i1} - p_{i2}$ = observed difference in proportions
w_i = $\dfrac{n_{i1}n_{i2}}{n_{i1} + n_{i2}}$: $w = \sum w_i$

Then we compute the weighted mean difference

$$\bar{d} = \frac{\sum w_i d_i}{w}$$

TABLE 9.
MORTALITY BY SEX OF DONOR AND SEVERITY OF DISEASE

Degree of disease	Sex of donor	Number of		Total	% deaths
		Deaths	Surv.		
None	M	2	21	23	8.7
	F	0	10	10	0.0
	Total	2	23	$33 = N_1$	$6.1 = p_1$
Mild	M	2	40	42	4.8
	F	0	18	18	0.0
	Total	2	58	$60 = N_2$	$3.3 = p_2$
Moderate	M	6	33	39	15.4
	F	0	10	10	0.0
	Total	6	43	$49 = N_3$	$12.2 = p_3$
Severe	M	17	16	33	51.5
	F	0	4	4	0.0
	Total	17	20	$37 = N_4$	$45.9 = p_4$

This has a standard error

$$\text{S.E.} = \frac{\sqrt{\sum w_i p_i q_i}}{w}$$

The test criterion is

$$\frac{\bar{d}}{\text{S.E.}} = \frac{\sum w_i d_i}{\sqrt{\sum w_i p_i q_i}}$$

This is referred to the tables of the normal distribution. As explained in the Appendix, which gives supporting reasons for this criterion, the criterion was constructed so that it would be powerful if the alternative hypothesis implies a constant difference on either the probit or the logit scale. The form of the criterion is not one that I would have selected intuitively, and the reader who feels the same way should consult the Appendix.

The test will be illustrated from data published by Diamond et al. (23). Erythroblastosis foetalis is a disease of newborn infants, some-

times fatal, caused by the presence in the blood of an *Rh*+ baby, of anti-*Rh* antibody transmitted by his *Rh*− mother. One form of treatment is an "exchange transfusion," in which as much as possible of the infant's blood is replaced by a donor's blood that is free of anti-*Rh* antibody. In 179 cases in which this treatment was used in a Boston hospital, the rather startling finding was made that there were no infant deaths out of 42 cases in which a female donor was used, but 27 infant deaths out of 137 cases in which a male donor was used. Since there seemed no *a priori* reason why there should be less hazard with female donors, a statistical investigation was made and reported in the reference.

One possibility was that male donors had been used primarily in the more severe cases. Consequently, the data were classified according to the stage of disease at birth, giving the four 2 × 2 tables shown in table 9.

The N's do not vary greatly, but the p's range from 3 to 46%. The combined test of significance is made from the supplementary data in table 10.

TABLE 10.
COMPUTATIONS FOR THE COMBINED TEST

Stage	d_i	$\hat{p}_i$	$\hat{p}_i\hat{q}_i$	$w_i = \dfrac{n_{i1}n_{i2}}{n_{i1} + n_{i2}}$
None	+ 8.7	6.1	573	7.0
Mild	+ 4.8	3.3	319	12.6
Moderate	+15.4	12.2	1071	8.0
Severe	+51.5	45.9	2483	3.6

$$\frac{d}{\text{S.E.}} = \frac{\sum w_i d_i}{\sqrt{\sum w_i \hat{p}_i \hat{q}_i}} = \frac{429.88}{\sqrt{25{,}537}} = 2.69$$

The significance probability is 0.0072. In data of this kind, a non-significant result would indicate that the surprising phenomenon can be explained by differences in the selection of cases. A significant result must be interpreted with caution, since conditions were not necessarily comparable for male and female donors even within cases of a given degree of severity. However, a significant result does encourage further study, e.g. by examining results in other hospitals.

The expectations in table 9 are so low that one might well doubt the validity of a normal approximation. At least, I did. However, the

exact distribution of $\sum w_i d_i$ can be worked out, with some labor, by writing down the probabilities of all possible configurations for each 2×2 table. The total numbers of configurations are 3, 3, 7 and 5, respectively, for the four tables, so that the total number of possible samples is 315, of which some have negligible probabilities. The value of the test criterion and the probability was worked out for each sample. The exact significance level was found to be 0.0095, as against the normal approximation of 0.0072. The degree of agreement is reassuring, considering the extreme smallness of the expectations. I am indebted to Mrs. Leah Barron for performing the computations.

9. THE EFFECT OF EXTRANEOUS VARIATION

Sometimes count data are subject to extraneous variation of a non-Poisson or non-binomial type, especially when the data are samples from an extensive population about which inferences are to be made. In these circumstances, an answer to the question that is really of interest frequently requires a test of significance which takes account of the extraneous variation. For this purpose, χ^2 tests are inappropriate, because they allow only for Poisson or binomial deviations from the null hypothesis. A few simple examples will be considered.

Suppose that we have a sample of data of the Poisson type and that the sample can be divided on some rational basis into g groups. If we wish to examine whether the mean varies from group to group, it was pointed out (section 3.3) that the variance X^2 can be divided into components as follows.

	d.f.
Between groups	$(g - 1)$
Within groups	$(N - g)$

(The latter component may be divided into a contribution from each group.) If the "Within groups" X^2 is statistically significant, this is a warning that the "Between groups" χ^2 test may be invalid. Further thought about the situation usually leads us to conclude that we do not want to declare that there are real differences between the true group means unless the observed group means differ by more than can be accounted for from the variation within groups. A more appropriate test is either an F-test of the "Between-groups" mean square against the "Within groups" mean square, or; if the observations are small and highly variable, an F-test in the square root scale. In fact, if there is reason to expect that the within-group variation will contain a non-Poisson component, use of an F-test without troubling to compute X^2 is a safer procedure. The same issue may arise in testing

whether p varies from group to group within a $2 \times N$ contingency table, where the χ^2 test may have to be replaced by an F-test based on the original p_i or on the equivalent angles. If the n_i vary substantially from one proportion to another, the problem of obtaining the most efficient type of F-test is discussed in (24).

Another example occurs in the combination of 2×2 contingency tables discussed in section 8. Suppose that the percentages of success under two procedures are compared repeatedly over a variety of conditions that are planned as a sample of some population of conditions about which we wish to draw conclusions. In this case the mean difference $\bar{d} = \bar{p}_1 - \bar{p}_2$ should be tested against the interaction with replications, instead of by the tests that were presented in section 8.

A thorough discussion of when not to use χ^2 tests in problems of this kind, and of the best alternatives, would be lengthy. It is hoped that the few words of warning given in this section will encourage critical thinking about the suitability of a χ^2 test.

10. CONCLUDING REMARKS

In presenting the examples in this paper, I have not attempted to give hard and fast rules as to when the supplementary tests should be used. The most useful principle is to think in advance about the way in which the data seem likely to depart from the null hypothesis. This often leads to the selection of a single test (e.g. the variance test) as the only one that appears appropriate. In other situations we may apply both the variance test (or some breakdown of it) and the goodness of fit test, on the grounds that the latter may reveal some types of departure that would not be discovered by the variance test.

When several tests are applied simultaneously to the same data, the chance that at least one of them will be significant is greater, and sometimes much greater, than the presumed 5% probability. This danger of misleading ourselves about the significance level is now widely recognized, and methods for avoiding it have been produced in some of the simpler problems. Although such methods need further development for the applications discussed in this paper, I believe that an awareness of the problem helps to prevent at least the worst distortions of the significance level.

In conclusion, I wish to thank Dr. Paul Meier for some useful suggestions.

APPENDIX

The purpose of this Appendix is to give some justification for the method proposed in section 8 for making a combined test of significance

of the difference between two proportions in a group of independent 2×2 tables. In the ith table, the observed proportions p_{i1} , p_{i2} are based on samples of size n_{i1} , n_{i2} , respectively, and the observed difference is $d_i = p_{i1} - p_{i2}$. The test criterion proposed is

$$\frac{\sum w_i d_i}{\sqrt{\sum w_i \hat{p}_i \hat{q}_i}}, \qquad \text{where } w_i = \frac{n_{i1} n_{i2}}{n_{i1} + n_{i2}}$$

and

$$\hat{p}_i = \frac{n_{i1} p_{i1} + n_{i2} p_{i2}}{n_{i1} + n_{i2}}$$

is the overall proportion from the margins of the ith table.

On the null hypothesis,

$$E(d_i) = 0; \qquad V(d_i) \doteq \hat{p}_i \hat{q}_i \left\{ \frac{1}{n_{i1}} + \frac{1}{n_{i2}} \right\} = \frac{\hat{p}_i \hat{q}_i}{w_i},$$

so that the test criterion is approximately normally distributed with mean 0 and s.d. 1. The problem is to show that the test criterion is sensitive in detecting departures from the null hypothesis.

The principal fact to be taken into account is that if the true p_i vary over a wide range, the true difference δ_i is unlikely to be constant on the observed (proportions) scale, but more likely to be constant on a probit or a logit scale. Suppose, for the moment, that we know the values of the δ_i on the alternative hypothesis, and consider the more general test criterion

$$Y = \frac{\sum W_i d_i}{\sqrt{\sum W_i^2 \hat{p}_i \hat{q}_i / w_i}}$$

where the W_i are any set of weights. This criterion reduces to the recommended criterion if $W_i = w_i$. It reduces to the sum of the individual X_i , divided by the square root of their number, if we put $W_i = 1/\sqrt{\hat{p}_i \hat{q}_i / w_i}$, and can ignore the contribution to the variance of Y arising from variation in the W_i . (This restriction is needed because we wish to treat the W_i as fixed weights).

On the null hypothesis, Y is approximately normally distributed with mean 0 and s.d. 1. On the alternative hypothesis,

$$E(\sum W_i d_i) = \sum W_i \delta_i$$

$$V(\sum W_i d_i) = \sum W_i^2 \left\{ \frac{p'_{i1} q'_{i1}}{n_{i1}} + \frac{p'_{i2} p'_{i2}}{n_{i2}} \right\}$$

where $'$ denotes a true proportion. Hence, if we can ignore the contribution to the variance of Y arising from its denominator (as we do in using χ^2 in a 2×2 table) we have

$$\frac{E(Y)}{\sigma(Y)} = \frac{\sum W_i \delta_i}{\sqrt{\sum W_i^2 \left\{ \frac{p'_{i1} q'_{i1}}{n_{i1}} + \frac{p'_{i2} q'_{i2}}{n_{i2}} \right\}}}$$

We will maximize the power of Y, on the alternative hypothesis, if we choose the W_i so as to maximize the ratio $E(Y)/\sigma(Y)$. By ordinary calculus methods, we find that we must have

$$W_i \propto \frac{\delta_i}{\frac{p'_{i1} q'_{i1}}{n_{i1}} + \frac{p'_{i2} q'_{i2}}{n_{i2}}}$$

Now if the δ_i are fairly small, so that p'_{i1} and p'_{i2} are not too far apart, the products $p'_{i1} q'_{i1}$ and $p'_{i2} q'_{i2}$ will both be approximately equal to $p'_i q'_i$, where p'_i is the mean of p'_{i1} and p'_{i2}. This gives, as an approximation to the best weights,

$$W_i \propto \frac{\delta_i}{p'_i q'_i \left(\frac{1}{n_{i1}} + \frac{1}{n_{i2}} \right)} = \frac{\delta_i w_i}{p'_i q'_i}$$

This result is not practically useful, since the δ_i are unknown. It happens, however, that if the true difference is constant on either the probit or the logit scale, the quantity $\delta_i / p'_i q'_i$ is close to constant over practically the whole range of p'_i from 0 to 100%.

In table 11, which illustrates this result, p'_{i1} has been given a range of values from 0 to 99%. To represent a constant effect on the probit scale, it is assumed that the true probit for the second sample is always 0.5 probit units higher than the probit of p'_{i1}. For the logit scale, the true logit for the second sample is taken as 0.8 logit units higher than the logit of p'_{i1}.

The sizes of the effects are roughly equal on the two scales for $p'_{i1} = 50\%$, where the alternative hypothesis gives about 69% successes, so that δ_i is about 19. For both the probit and logit cases, δ_i varies widely from this value as p'_{i1} varies, but the variation in δ_i is closely matched by that in $p'_i q'_i$ so that the ratio is practically constant for the logit case and varies only slightly for the probit case. If the δ_i are made smaller, the constancy of the ratio is improved; if larger, the ratio varies more.

The conclusion from this argument is that if effects are constant, for varying levels of p'_{i1}, on either the probit or logit scale, the choice of $W_i = w_i$ gives a test criterion that should be close to the optimum in power.

TABLE 11.
CHECK ON CONSTANCY OF $\delta_i/p'_iq'_i$

p'_{i1}(%)	Constant probit effect			Constant logit effect		
	δ_i	$p'_iq'_i$	$10^4\delta_i/p'_iq'_i$	δ_i	$p'_iq'_i$	$10^4\delta_i/p'_iq'_i$
1	2.38	214	111	1.20	157	76
5	7.61	803	95	5.49	714	77
10	11.71	1334	88	9.83	1269	77
30	19.04	2390	80	18.82	2388	79
50	19.15	2408	80	19.00	2410	79
70	14.71	1751	84	13.85	1775	78
90	6.26	640	98	5.24	683	77
95	3.40	319	107	2.69	352	76
99	0.76	62	122	0.55	72	76

REFERENCES

(1) D. Lewis and C. J. Burke, "The use and misuse of chi-square test," *Psych. Bull.*, *46* (1949), 433–498.

(2) P. V. Sukhatme, "On the distribution of χ^2 in small samples of the Poisson series," *Jour. Roy. Stat. Soc. Suppl.*, *5* (1938), 75–79.

(3) J. Neyman and E. S. Pearson, "Further notes on the χ^2 distribution," *Biometrika*, *22* (1931), 298–305.

(4) W. G. Cochran, "The χ^2 correction for continuity," *Iowa State Coll. Jour. Sci.*, *16* (1942), 421–436.

(5) W. G. Cochran, "The χ^2 distribution for the binomial and Poisson series, with small expectations," *Annals of Eugenics*, *7* (1936), 207–217.

(6) E. Fix, "Tables of the Noncentral χ^2," *Univ. California Publ. Stat.*, *1* (1949), 15–19.

(7) W. G. Cochran, "The χ^2 test of goodness of fit," *Annals of Math. Stat.*, *23* (1952), 315–345.

(8) D. Mainland, "Statistical methods in medical research," *Canadian Jour. Res.*, *E*, *26* (1948), 1–166.

(9) G. H. Freeman and J. H. Halton, "Note on an exact treatment of contingency, goodness of fit and other problems of significance," *Biometrika*, *38* (1951), 141–149.

(10) J. B. S. Haldane, The mean and variance of χ^2 when used as a test of homogeneity, when expectations are small," *Biometrika*, *31* (1939), 346–355.

(11) H. O. Lancaster, "The derivation and partition of χ^2 in certain discrete distributions," *Biometrika*, *36* (1949), 117–129.

(12) W. G. Cochran, "Statistical analysis of field counts of diseased plants," *Jour. Roy. Stat. Soc. Suppl.*, *3* (1936), 49–67.
(13) D. P. Murphy, *Heredity in uterine cancer*, Harvard University Press, 1952.
(14) F. Yates, "The analysis of contingency tables with groupings based on quantitative characters," *Biometrika*, *35* (1948), 176–181.
(15) J. O. Irwin, "A note on the subdivision of χ^2 into components," *Biometrika*, *36* (1949), 130–134.
(16) A. W. Kimball, "Short-cut formulas for the exact partition of χ^2 in contingency tables," *Biometrics*, *10* (1954), 452–458.
(17) C. P. Winsor, "Factorial analysis of a multiple dichotomy," *Human Biology*, *20* (1948), 195–204.
(18) F. J. Anscombe, "Sampling theory of the negative binomial and logarithmic distributions," *Biometrika*, *37* (1950), 358–382.
(19) C. I. Bliss, "Fitting the negative binomial distribution to biological data," *Biometrics*, *9* (1953), 176–196.
(20) J. Berkson, "A note on the chi-square test, the Poisson and the binomial," *Jour. Amer. Stat. Ass.*, *35* (1940), 362–367.
(21) J. Berkson, "Some difficulties of interpretation encountered in the application of the chi-square test," *Jour. Amer. Stat. Ass.*, *33* (1938), 526–536.
(22) G. V. Dyke and H. D. Patterson, "Analysis of factorial arrangements when the data are proportions," *Biometrics*, *8* (1952), 1–12.
(23) Fred H. Allen, Jr., Louis K. Diamond and Joseph B. Watrous, Jr., "Erythroblastosis fetalis. V. The value of blood from female donors for exchange transfusion," *The New England Jour. of Med.*, *241* (1949), 799–806.
(24) W. G. Cochran, "Analysis of variance for percentages based on unequal numbers," *Jour. Amer. Stat. Assoc.*, *38* (1943), 287–301.

SHORT-CUT FORMULAS FOR THE EXACT PARTITION OF χ^2 IN CONTINGENCY TABLES*

A. W. KIMBALL

Oak Ridge National Laboratory

1. INTRODUCTION AND SUMMARY

Recently, Lancaster (1949, 1950) and Irwin (1949) showed how χ^2 may be partitioned exactly into single degrees of freedom when expected frequencies are estimated from the data. In particular they considered the case of an $r \times s$ (r rows, s columns) contingency table in which the expected frequencies are computed from the marginal totals. Their work suggests several different methods of carrying out the necessary computations, some of which they have illustrated. For tables of any order, their contingency table formulas may be reduced to surprisingly simple and useful forms which effect a considerable saving of computational labor. The short-cut formulas are given here without proof since the reduction involves only straightforward algebra. Since this paper is intended primarily as a simplification of the Lancaster-Irwin formulas, their notation has been changed slightly in the hope that computers will find the formulas easier to read. The method is applied to a radiation experiment with grasshopper neuroblasts.

2. THE METHODS AND FORMULAS

Let the observed frequencies and marginal totals of an $r \times s$ contingency table be

a_1	a_2	a_3	$\cdots$	a_s	A
b_1	b_2	b_3	$\cdots$	b_s	B
c_1	c_2	c_3	$\cdots$	c_s	C
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
z_1	z_2	z_3	$\cdots$	z_s	Z
n_1	n_2	n_3	$\cdots$	n_s	N

*Presented at the American Statistical Association Annual Meeting in Montreal, 1954, in a session held jointly with The Biometric Society (*ENAR*).

where there are r letters inclusive between a and z. The χ^2 as usually computed from this table has $(r - 1)(s - 1)$ degrees of freedom. The first step in subdivision is to find $(r - 1)(s - 1)$ fourfold tables each of which contains certain subtotals of the observed frequencies. These tables are constructed most simply by forming, for each frequency in the last $(r - 1)$ rows and $(s - 1)$ columns, the fourfold table which contains this frequency in the lower right hand cell. The remaining three cells are composed of the frequencies above and to the left of the frequency chosen for the lower right hand cell. A few examples should serve to make the pattern clear. In a 2×3 table, the two fourfold tables needed correspond to the frequencies b_2 and b_3 and are

$$\begin{array}{cc|cc} a_1 & a_2 & a_1 & a_2 & a_3 \\ \hline b_1 & b_2 & b_1 & b_2 & b_3 \end{array}.$$

The frequencies in any one cell should be added, e.g., $(a_1 + a_2)$ and $(b_1 + b_2)$, but the pattern is made clearer if the tables are presented as above. Likewise, for a 3×4 table the six fourfold tables are

$$\begin{array}{cc|cc|c} a_1 & a_2 & a_1 & a_2 & a_3 \\ \hline b_1 & b_2 & b_1 & b_2 & b_3 \\ \hline \end{array} \qquad \begin{array}{ccc|c} a_1 & a_2 & a_3 & a_4 \\ \hline b_1 & b_2 & b_3 & b_4 \\ \hline \end{array}$$

$$\begin{array}{cc|cc|c} a_1 & a_2 & a_1 & a_2 & a_3 \\ b_1 & b_2 & b_1 & b_2 & b_3 \\ \hline c_1 & c_2 & c_1 & c_2 & c_3 \end{array} \qquad \begin{array}{ccc|c} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ \hline c_1 & c_2 & c_3 & c_4 \end{array}$$

Note that the elements of the last two rows and last three columns of the 3×4 table appear successively in the lower right hand cells of the fourfold tables. In practice these fourfold tables may not have to be written down separately, but the reader should bear in mind that each of the single degree of freedom chi-squares is associated with one of the tables.

One of the methods apparently favored by Lancaster and Irwin involves the construction and subsequent multiplication of three Helmert-type matrices. The method suggested here seems to require much less effort. The short-cut formulas which appear to be most useful are:

$$2 \times 3 \qquad \chi_1^2 = \frac{N^2[a_1b_2 - a_2b_1]^2}{ABn_1n_2(n_1 + n_2)}$$

$$\chi_2^2 = \frac{N[b_3(a_1 + a_2) - a_3(b_1 + b_2)]^2}{ABn_3(n_1 + n_2)}$$

2×4 $$\chi_1^2 = \frac{N^2[a_1b_2 - a_2b_1]^2}{ABn_1n_2(n_1 + n_2)}$$

$$\chi_2^2 = \frac{N^2[b_3(a_1 + a_2) - a_3(b_1 + b_2)]^2}{ABn_3(n_1 + n_2)(n_1 + n_2 + n_3)}$$

$$\chi_3^2 = \frac{N[b_4(a_1 + a_2 + a_3) - a_4(b_1 + b_2 + b_3)]^2}{ABn_4(n_1 + n_2 + n_3)}$$

$2 \times s$ $$\chi_t^2 = \frac{N^2[b_{t+1}S_t^{(a)} - a_{t+1}S_t^{(b)}]^2}{ABn_{t+1}S_t^{(n)}S_{t+1}^{(n)}} \qquad t = 1, \cdots, s-1$$

where

$$S_t^{(a)} = \sum_{i=1}^{t} a_i , \qquad S_t^{(b)} = \sum_{i=1}^{t} b_i , \qquad S_t^{(n)} = \sum_{i=1}^{t} n_i .$$

3×3 $$\chi_1^2 = \frac{N[B(n_2a_1 - n_1a_2) - A(n_2b_1 - n_1b_2)]^2}{ABn_1n_2(A + B)(n_1 + n_2)}$$

$$\chi_2^2 = \frac{N^2[b_3(a_1 + a_2) - a_3(b_1 + b_2)]^2}{ABn_3(A + B)(n_1 + n_2)}$$

$$\chi_3^2 = \frac{N^2[c_2(a_1 + b_1) - c_1(a_2 + b_2)]^2}{Cn_1n_2(A + B)(n_1 + n_2)}$$

$$\chi_4^2 = \frac{N[c_3(a_1 + a_2 + b_1 + b_2) - (a_3 + b_3)(c_1 + c_2)]^2}{Cn_3(A + B)(n_1 + n_2)}$$

In a general $r \times s$ table, the single degree of freedom corresponding to the element k_m may be obtained from the fourfold table

a_1	a_2	$\cdots$	a_{m-1}	a_m
b_1	b_2	$\cdots$	b_{m-1}	b_m
$\vdots$	$\vdots$		$\vdots$	$\vdots$
j_1	j_2	$\cdots$	j_{m-1}	j_m
k_1	k_2	$\cdots$	k_{m-1}	k_m

by means of the formula

$$\chi_{k-}^2 = \frac{\left\{\begin{array}{l} Nn_m[K(S_{m-1}^{(a)} + S_{m-1}^{(b)} + \cdots + S_{m-1}^{(j)}) - (A+B+\cdots+J)S_{m-1}^{(k)}] \\ -S_{m-1}^{(n)}[K(a_m + b_m + \cdots + j_m) - (A+B+\cdots+J)k_m] \end{array}\right\}^2}{n_mS_{m-1}^{(n)}S_m^{(n)}K(A+B+\cdots+J)(A+B+\cdots+J+K)} .$$

Each of the formulas for the single degree of freedom chi-squares reduces to one like the familiar formula for a 2×2 table. Only the observed frequencies and marginal totals are involved so that expected frequencies never have to be computed. In this respect computation of the single degree of freedom chi-squares from an $r \times s$ table requires no more effort than the computation of chi-squares from $(r-1)(s-1)$ 2×2 tables.

Lancaster (1950) gives the following 3×3 table to illustrate the method:

3	4	15	22
2	8	5	15
2	6	5	13
7	18	25	50

Using the short-cut formulas, we find:

$$\chi_1^2 = \frac{50[15(18 \times 3 - 7 \times 4) - 22(18 \times 2 - 7 \times 8)]^2}{(22)(15)(7)(18)(37)(25)} = 0.895571$$

$$\chi_2^2 = \frac{50^2[5 \times 7 - 15 \times 10]^2}{(22)(15)(25)(37)(25)} = 4.332514$$

$$\chi_3^2 = \frac{50^2[6 \times 5 - 2 \times 12]^2}{(13)(7)(18)(37)(25)} = 0.059400$$

$$\chi_4^2 = \frac{50[5 \times 17 - 8 \times 20]^2}{(13)(25)(37)(25)} = 0.935551.$$

These values agree with Lancaster's and, of course, sum to 6.22304 which is the total chi-square with 4 d.f.

3. A BIOLOGICAL APPLICATION

An interesting application of this technique was found in the interpretation of some radiation experiments. Dr. Mary Esther Gaulden, of the Biology Division, Oak Ridge National Laboratory, has been comparing the effects of X rays and beta rays on mitotic rates in grasshopper neuroblasts (*Chortophaga viridifasciata*). In one series of experiments, embryos from the same egg case were divided into three groups. One group served as controls and the other two groups were exposed to physically equivalent doses of X and beta radiation. After irradiation, approximately equal numbers of cells in the three groups were examined periodically for 209 minutes, the average length of the

TABLE 1

PARTITIONING OF CHI-SQUARE IN SOME RADIATION EXPERIMENTS.
(Data from Dr. Mary Esther Gaulden)

Experiment Number		Observed Frequencies				Chi-Squares		
		Group			Totals	χ^2_T	χ^2_1	χ^2_2
		X-ray	Beta	Control				
1	*a*	3	4	12	19	7.133	0.262	6.871
	b	15	12	10	37			
	n	18	16	22	56			
2	*a*	3	3	12	18	11.752	0.000	11.752
	b	12	12	5	29			
	n	15	15	17	47			
3	*a*	5	2	11	18	10.026	0.730	9.296
	b	17	17	9	43			
	n	22	19	20	61			
4	*a*	3	3	17	23	29.309	0.003	29.306
	b	16	17	2	35			
	n	19	20	19	58			
Pooled	*a*	14	12	52	78	52.510	0.050	52.460
	b	60	58	26	144			
	n	74	70	78	222			

a = Number of cells reaching mid-mitosis
b = Number of cells not reaching mid-mitosis
n = Total number of cells

mitotic cycle in normal cells. In each group the number of cells passing through mid-mitosis (prometaphase, metaphase and anaphase) was noted. Since irradiation tends to reduce the mitotic rate, the proportion of cells which reach mid-mitosis is an index of the amount of radiation damage.

Table 1 contains the results of four such experiments. The data from each experiment form a 2×3 contingency table from which a

total chi-square (χ_T^2) with two d.f. may be computed. If neither radiation treatment has any effect, the χ_T^2 should exceed 5.991 only about 5 per cent of the time. Since all four experiments yielded significant χ_T^2's, we must conclude that irradiation does reduce the mitotic rate.

Dr. Gaulden, however, was also interested in a comparison of the effects of the two different types of radiation. A test for this comparison can be made easily and accurately by partitioning χ_T^2 into appropriate individual d.f. The partitioning formulas given previously may be applied directly to the data as they appear in Table 1. For example, from the first experiment, the fourfold table,

3	4
15	12

,

provides a comparison of X and beta radiations. The single d.f. chi-square associated with this table is

$$\chi_1^2 = \frac{56^2(3 \times 12 - 4 \times 15)^2}{19 \times 37 \times 18 \times 16 \times 34} = 0.262.$$

The other d.f. is associated with the fourfold table,

7	12
27	10

,

and provides a comparison of the control group with both irradiated groups by means of the single d.f. chi-square

$$\chi_2^2 = \frac{56(7 \times 10 - 12 \times 27)^2}{19 \times 37 \times 34 \times 22} = 6.871.$$

Since this method provides an exact partitioning of χ_T^2,

$$\chi_T^2 = \chi_1^2 + \chi_2^2 = 0.262 + 6.871 = 7.133.$$

The 5 per cent point of the chi-square distribution for one d.f. is 3.841; thus, there is no evidence of any difference between the two types of radiation.

In this problem the partitioning has even another advantage. The original test of overall radiation effect was made with $\chi_T^2 = 7.133$ which exceeded the 5 per cent point for two d.f. only slightly. On the other

hand, $\chi_2^2 = 6.871$ is almost twice the 5 per cent point for one d.f. and obviously provides a more sensitive test of the overall radiation effect. The difference in sensitivity is attributable directly to the fact that χ_T^2 is, so to speak, diluted down by the absence of a difference between X and beta radiation. In χ_2^2, this dilution has been removed.

We have considered only one experiment in detail. The results for the other three experiments are essentially the same and are given in Table 1. If the experiments can be taken as identical except for sampling fluctuations, a further increase in sensitivity may be achieved by pooling the data from all four experiments and partitioning the chi-square obtained from it. This has also been done in Table 1. If the experiments differ because of environmental conditions, source of material or any other factors, pooling is not permissible but the χ_1^2's and χ_2^2's from the individual experiments may be added and the resulting values compared with significance levels of chi-square for 4 d.f. In general, these tests will not be as sensitive as tests based on the pooled data. For example from Table 1, the χ_2^2 from the pooled data is 52.460 and has one d.f., whereas the sum of the individual χ_2^2's is 57.225 and has four d.f. The former is significant at a much lower level of probability than is the latter. In a sense this may be interpreted as the price one must pay for failure to maintain uniform conditions.

The author is indebted to Dr. Gaulden for permission to use her data and to George J. Atta for checking the computations.

REFERENCES

[1] Irwin, J. O. "A note on the subdivision of χ^2 into components", *Biometrika*, 36 (1949), 130–134.

[2] Lancaster, H. O. "The derivation and partition of χ^2 in certain discrete distributions", *Biometrika*, 36 (1949), 117–129.

[3] Lancaster, H. O. "The exact partition of χ^2 and its application to the problem of pooling of small expectations", *Biometrika*, 37 (1950), 267–270.

A NOTE ON THE χ^2 TEST APPLIED TO EPIDEMIC CHAINS

JOYCE ALMOND
(University of Manchester, England)

1. The flow of a disease in a population is often represented by a mathematical model based on a set of assumptions about the spread of the disease. Dr. Helen Abbey in her paper "An Examination of the Reed-Frost Theory of Epidemics" attempted to determine the fit of such a model to actual observed epidemics, by means of a χ^2 test.

The particular model under investigation was the Reed-Frost model, which is based on assumptions stated in Abbey's paper. The model is a discrete one, the length of duration of the epidemic being divided into equal time intervals of suitable length. The epidemic is supposed to take place in a closed population. Let p be the probability of contact between any two specified individuals, where "contact" is such that if it occurs between an infected person and a susceptible, another infectious case is produced. Once an individual becomes a case, he passes out of observation for the rest of the epidemic, being neither a case nor a susceptible for future time periods. If c_t and n_t are the number of cases and susceptibles respectively, observed in the time interval t*, then

$$n_t = c_{t+1} + n_{t+1} ,$$

and the probability of c_{t+1} cases in the $(t + 1)$th interval is

$$p(c_{t+1}) = \frac{n_t!}{c_{t+1}!n_{t+1}!}(1 - q^{c_t})^{c_{t+1}}(q^{c_t})^{n_{t+1}} \tag{1}$$

where $q = 1 - p$. The *expected* numbers of cases and susceptibles for the time period $(t + 1)$ are given by

$$\begin{aligned} E(c_{t+1}) &= n_t(1 - q^{c_t}) \\ E(n_{t+1}) &= n_t - E(c_{t+1}) \end{aligned} \tag{2}$$

where, as above, n_t and c_t are the *observed* numbers of susceptibles and cases for the time period t.

The diseases which are most likely to satisfy the conditions of the model are measles, German measles and chicken pox, when an epidemic occurs in an institution, which can be considered to form a closed population approximately. Accordingly the model was tested against such epidemics for which data can be obtained.

*I have changed Abbey's notation here, using n_t instead of s_t for number of susceptibles at time t.

In applying the model to observed data, Abbey described how each epidemic was divided into intervals of the length of the incubation period of the disease, such that as far as possible there was a clustering of cases in the middle of each interval. The Reed-Frost model was then applied to the total numbers of cases and susceptibles in each interval.

An estimate of the contact rate had to be made for each epidemic. A single constant contact rate for the whole epidemic was obtained by maximum likelihood methods.

$$\text{likelihood} = \prod_{r=1}^{k} \frac{n_r!}{c_{r+1}!n_{r+1}!}(1 - q^{c_r})^{c_{r+1}}(q^{c_r})^{n_{r+1}}$$

The log likelihood, denoted by L, was then maximised with respect to q in order to obtain the maximum likelihood estimate $\hat{q}$.

2. A χ^2 Goodness of Fit Test was then carried out by forming

$$\chi^2 = \sum_{r=1}^{k} \frac{[Oc_r - E(c_r)]^2}{E(c_r)} + \sum_{r=1}^{k} \frac{[On_r - E(n_r)]^2}{E(n_r)}. \qquad (3)$$

The classical χ^2 Goodness of Fit Test is formed under the assumption that the number of observations falling into a particular group is independent of the number falling into any other group, apart from the linear restriction, if, as usual, the total number is given. Here, the time periods correspond to groups, and the number of cases (or number of susceptibles) corresponds to the number of observations. In this case, in the actual observed epidemic, the number of cases in the period $(t + 1)$ will be expected to depend on the number in the previous period t. In the theoretical model, the probability of c_{t+1} cases in the period $(t + 1)$ is defined *given* that there are c_t cases in the period t, so that the property of independence no longer holds. In particular, if the number of cases (or the number of susceptibles) is zero in the time period t, the epidemic will automatically cease at time t, and the number of cases in all future time periods in this epidemic will be zero. So the appearance of any "observations" in group $(t + 1)$ is dependent on the number of observations in the previous group being non-zero.

The expected number of cases in time period $(t + 1)$, $E(c_{t+1})$, is calculated from the value of the observed number of cases in the period t. Thus c_t is considered first as a random variable in the time period t, and then as a known constant for the following period. The same is true of course when considering the numbers of susceptibles.

In order to investigate the form of the distribution of the above "χ^2", consider first the case when the contact rate is assumed known

and does not have to be estimated from the data. For simplicity consider the χ^2 which would be formed from the first two stages of the epidemic only. This would give

$$\chi^2 = \frac{[c_1 - n_0(1 - q^{c_0})]^2}{n_0(1 - q^{c_0})} + \frac{[c_2 - (n_0 - c_1)(1 - q^{c_1})]^2}{(n_0 - c_1)(1 - q^{c_1})}$$
$$+ \frac{[c_1 - n_0(1 - q^{c_0})]^2}{n_0 q^{c_0}} + \frac{[c_2 - (n_0 - c_1)(1 - q^{c_1})]^2}{(n_0 - c_1)q^{c_1}} \qquad (4)$$
$$= \frac{[c_1 - n_0(1 - q^{c_0})]^2}{n_0 q^{c_0}(1 - q^{c_0})} + \frac{[c_2 - (n_0 - c_1)(1 - q^{c_1})]^2}{(n_0 - c_1)q^{c_1}(1 - q^{c_1})}$$

where c_0 , n_0 the initial numbers of cases and susceptibles are assumed known. Here the number of cases at the first stage c_1 behaves as a random variable in the first term, and as a fixed quantity in the second term, which corresponds to the second stage of the epidemic.

If c_1 or $n_0 - c_1 = 0$, as noted above $c_2 = 0$, and the term in χ^2 corresponding to the second time period will vanish, as would be expected. In this case the epidemic "dies out" at the first stage.

Probability that $c_1 = 0$ is $q^{c_0 n_0}$

Probability that $c_1 = n_0$ is $(1 - q^{c_0})^{n_0}$

so the probability of an epidemic breaking out is

$$1 - q^{c_0 n_0} - (1 - q^{c_0})^{n_0}$$

Strictly, this means that when the epidemic extends to at least 2 stages, so that χ^2 has at least two terms in it, the probability of c_1 is altered from

$$p(c_1) = \frac{n_0!}{(n_0 - c_1)!c_1!}(1 - q^{c_0})^{c_1}(q^{c_0})^{n_0 - c_1}$$

to

$$p(c_1 \mid 0 < c_1 < n_0) = \frac{1}{[1 - q^{c_0 n_0} - (1 - q^{c_0})^{n_0}]}$$
$$\cdot \frac{n_0!}{(n_0 - c_1)!c_1!}(1 - q^{c_0})^{c_1}(q^{c_0})^{n_0 - c_1}$$

Considering n_0 large, the term $(1 - q^{c_0})^{n_0}$, which is the probability of all the n_0 susceptibles being infected at the first stage, will certainly be negligible here, and the term $q^{c_0 n_0}$ will also be very small compared with unity.

The case $c_1 = 0$ (or $n_0 = c_1$) does not seem to have been considered by Abbey, as she has taken only those epidemics which do not die out immediately after the initial case or cases.

So taking

$$\chi^2 = \frac{[c_1 - n_0(1 - q^{c_0})]^2}{n_0(1 - q^{c_0})q^{c_0}} + \frac{[c_2 - (n_0 - c_1)(1 - q^{c_1})]^2}{(n_0 - c_1)(1 - q^{c_1})q^{c_1}}$$

and assuming that approximately

$$p(c_1) = \frac{n_0!}{c_1!(n_0 - c_1)!}(1 - q^{c_0})^{c_1}(q^{c_0})^{n_0 - c_1}$$

we find

$$E(\chi^2) = 1 + 1 = 2 \tag{5}$$

averaging first over c_2 with c_1 fixed.
To find the variance of χ^2 form

$$\chi^4 = \frac{[c_1 - n_0(1 - q^{c_0})]^4}{[n_0(1 - q^{c_0})q^{c_0}]^2} + \frac{[c_2 - (n_0 - c_1)(1 - q^{c_1})]^4}{[(n_0 - c_1)(1 - q^{c_1})q^{c_1}]^2} + \frac{2[c_1 - n_0(1 - q^{c_0})]^2[c_2 - (n_0 - c_1)(1 - q^{c_1})]^2}{n_0(1 - q^{c_0})q^{c_0}(n_0 - c_1)(1 - q^{c_1})q^{c_1}} \tag{6}$$

$$= X + Y + Z \quad \text{(say)}.$$

$$E\{Z\} = 2 \text{ (as before, averaging first over } c_2 \text{ with } c_1 \text{ fixed.)}$$

$$E\{X\} = 3 + \frac{1 - 6q^{c_0}(1 - q^{c_0})}{n_0(1 - q^{c_0})q^{c_0}}$$

and similarly

$$E\{Y\} = 3 + E_{c_1}\left\{\frac{1}{(n_0 - c_1)(1 - q^{c_1})q^{c_1}}\right\} - 6E_{c_1}\left\{\frac{1}{n_0 - c_1}\right\}$$

Denoting $E_{c_1}\{n_0 - c_1\}$ by $\overline{n_0 - c_1}$, we have

$$E_{c_1}\left\{\frac{1}{n_0 - c_1}\right\} = E_{c_1}\left\{\frac{1}{\overline{n_0 - c_1} + (n_0 - c_1) - \overline{n_0 - c_1}}\right\}$$

$$= E_{c_1}\left\{\frac{1}{(\overline{n_0 - c_1})} - \frac{(n_0 - c_1 - \overline{n_0 - c_1})}{(\overline{n_0 - c_1})^2} + \frac{(n_0 - c_1 - \overline{n_0 - c_1})^2}{(\overline{n_0 - c_1})^3} + \cdots\right\}$$

$$= \frac{1}{\overline{n_0 - c_1}} + \frac{\sigma^2(n_0 - c_1)}{(\overline{n_0 - c_1})^3} \cdots$$

$$= \frac{1}{n_0 q^{c_0}}\left\{1 + \frac{1 - q^{c_0}}{n_0 q^{c_0}} + \cdots\right\}.$$

Since

$$(n_0 - c_1)(1 - q^{c_1})q^{c_1} = (n_0 - \bar{c_1})(1 - q^{c_1})q^{c_1}\left\{1 - \frac{c_1 - \bar{c_1}}{n_0 - \bar{c_1}}\right\}$$

$$= (n_0 - \bar{c_1})(1 - q^{c_1})q^{c_1}\left\{1 + \frac{0\sqrt{(n_0(1 - q^{c_0})q^{c_0})}}{n_0 q^{c_0}}\right\}$$

$$= (n_0 - \bar{c_1})(1 - q^{c_1})q^{c_1}\left\{1 + 0\sqrt{\frac{(1 - q^{c_0})}{n_0 q^{c_0}}}\right\},$$

$$E_{c_1}\left\{\frac{1}{(n_0 - c_1)(1 - q^{c_1})q^{c_1}}\right\} \sim \frac{1}{(n_0 - \bar{c_1})} E_{c_1}\left\{\frac{1}{(1 - q^{c_1})q^{c_1}}\right\}$$

but writing $q^{c_1} - q^{2c_1} = \overline{q^{c_1} - q^{2c_1}} + \delta(q^{c_1} - q^{2c_1})$, we obtain further

$$E_{c_1}\left\{\frac{1}{q^{c_1}(1 - q^{c_1})}\right\} = \frac{1}{\overline{q^{c_1} - q^{2c_1}}}\left\{1 + \frac{\sigma^2(\delta)}{(\overline{q^{c_1} - q^{2c_1}})^2} \cdots\right\}$$

Now

$$E_{c_1}\{q^{c_1 r}\} = \sum_{c_1=0}^{n_0} \frac{n_0!}{c_1!(n_0 - c_1)!}(1 - q^{c_0})^{c_1}(q^{c_0})^{n_0 - c_1} q^{c_1 r}$$

$$= [(1 - q^{c_0})q^r + q^{c_0}]^{n_0}$$

but

$$(1 - q^{c_0})q^r + q^{c_0} = \{1 - (1 - p)^{c_0}\}(1 - p)^r + (1 - p)^{c_0}$$

$$= (1 - p)^{c_0} + (1 - p)^r - (1 - p)^{c_0 + r}$$

$$= 1 - rc_0p^2 + \tfrac{1}{2}rc_0p^3[r + c_0 - 2] + \cdots$$

So for p small, (not much greater than $1/\sqrt{n_0}$), and r, $c_0 \sim 1$,

$$[(1 - q^{c_0})q^r + q^{c_0}]^{n_0} = [1 - rc_0p^2 + \tfrac{1}{2}p^3c_0r(r + c_0 - 2)]^{n_0}$$

which, for n_0 large, $\sim \exp\{-n_0c_0rp^2[1 - \frac{1}{2}p(r + c_0 - 2)]\}$ to the same order of approximation. Hence

$$E_{c_1}\left\{\frac{1}{(n_0 - \bar{c_1})(1 - q^{c_1})q^{c_1}}\right\} = \frac{1}{n_0 q^{c_0} A}\left[1 + \frac{B}{A^2}\right]$$

where

$$\left.\begin{aligned} A &= e^{-n_0c_0p^2[1 - \frac{1}{2}p(c_0 - 1)]} - e^{-2n_0c_0p^2[1 - \frac{1}{2}pc_0]} \\ \text{and} \\ B &= e^{-2n_0c_0p^2[1 - \frac{1}{2}pc_0]} - 2e^{-3n_0c_0p^2[1 - \frac{1}{2}p(c_0 + 1)]} \\ &\quad + e^{-4n_0c_0p^2[1 - \frac{1}{2}p(c_0 + 2)]} - e^{-2n_0c_0p^2[1 - \frac{1}{2}p(c_0 - 1)]} \\ &\quad + 2e^{-n_0c_0p^2[3 - \frac{1}{2}p(3c_0 - 1)]} - e^{-4n_0c_0p^2[1 - \frac{1}{2}pc_0]} \end{aligned}\right\} \quad (7)$$

Hence

$$\sigma^2(\chi^2) = 4 + \frac{1 - 6q^{c_0}(1 - q^{c_0})}{n_0(1 - q^{c_0})q^{c_0}} - \frac{6}{n_0 q^{c_0}}\left\{1 + \frac{(1 - q^{c_0})}{n_0 q^{c_0}} + \cdots\right\} + \left\{1 + 0\sqrt{\frac{(1 - q^{c_0})}{n_0 q^{c_0}}}\right\}\frac{1}{n_0 q^{c_0} A}\left[1 + \frac{B}{A^2}\right]. \tag{8}$$

So that it appears that under certain conditions, namely for initial number of susceptibles n_0 very large and probability of contact p small, the first two moments of this "χ^2" will be approximately equal to those of the classical χ^2 for two classes.

The extension to k stages is indicated below. In this case

$$\chi^2 = \frac{[c_1 - n_0(1 - q^{c_0})]^2}{n_0(1 - q^{c_0})q^{c_0}} + \cdots + \frac{[c_{k-1} - n_{k-2}(1 - q^{c_{k-2}})]^2}{n_{k-2}(1 - q^{c_{k-2}})q^{c_{k-2}}} + \frac{[c_k - n_{k-1}(1 - q^{c_{k-1}})]^2}{n_{k-1}(1 - q^{c_{k-1}})q^{c_{k-1}}} \tag{9}$$

provided $c_1 , c_2 \cdots c_{k-1} \neq 0$, $n_1 , n_2 \cdots n_{k-1} \neq 0$.
Strictly the moment generating function of c_r given $c_1 \cdots c_{r-1}$ $(1 \leq r \leq k - 1)$ will be given by:

$$M'_{c_r}(\theta) = \frac{[1 + (1 - q^{c_{r-1}})(e^\theta - 1)]^{n_{r-1}} - q^{c_{r-1}n_{r-1}} - (1 - q^{c_{r-1}})^{n_{r-1}}e^{\theta n_{r-1}}}{1 - q^{c_{r-1}n_{r-1}} - (1 - q^{c_{r-1}})^{n_{r-1}}}$$

so

$$E'_{c_r}(c_r^s) = \frac{1}{1 - q^{c_{r-1}n_{r-1}} - (1 - q^{c_{r-1}})^{n_{r-1}}}\left[\left\{\text{coeff of } \frac{\theta^s}{s!} \text{ in } [1 + (1 - q^{c_{r-1}})(e^\theta - 1)]^{n_{r-1}}\right\} - (1 - q^{c_{r-1}})^{n_{r-1}}(n_{r-1})^s\right],$$

where the dash indicates expectation conditional on $0 < c_r < n_{r-1}$. As in the case of 2 stages only, for n_{r-1} sufficiently large, the smaller moments will be approximately equal to the usual binomial moments, e.g.

$$E'_{c_r}\{c_r\} = \frac{n_{r-1}(1 - q^{c_{r-1}}) - n_{r-1}(1 - q^{c_{r-1}})^{n_{r-1}}}{1 - q^{c_{r-1}n_{r-1}} - (1 - q^{c_{r-1}})^{n_{r-1}}}$$

$$= n_{r-1}(1 - q^{c_{r-1}})\frac{\{1 - (1 - q^{c_{r-1}})^{n_{r-1}-1}\}}{1 - q^{c_{r-1}n_{r-1}} - (1 - q^{c_{r-1}})^{n_{r-1}}} \sim n_{r-1}(1 - q^{c_{r-1}})$$

and

$$E'_{c_r}\{c_r^2\} = \frac{n_{r-1}(1 - q^{c_{r-1}})q^{c_{r-1}} + n_{r-1}^2(1 - q^{c_{r-1}})^2 - n_{r-1}^2(1 - q^{c_{r-1}})^{n_{r-1}}}{1 - q^{c_{r-1}n_{r-1}} - (1 - q^{c_{r-1}})^{n_{r-1}}}$$

$$= \frac{n_{r-1}(1 - q^{c_{r-1}})q^{c_{r-1}}}{1 - q^{c_{r-1}n_{r-1}} - (1 - q^{c_{r-1}})^{n_{r-1}}}$$

$$+ \frac{n_{r-1}^2(1 - q^{c_{r-1}})^2\{1 - (1 - q^{c_{r-1}})^{n_{r-1}-2}\}}{1 - q^{c_{r-1}n_{r-1}} - (1 - q^{c_{r-1}})^{n_{r-1}}}$$

$$\sim n_{r-1}(1 - q^{c_{r-1}})q^{c_{r-1}} + n_{r-1}^2(1 - q^{c_{r-1}})^2.$$

Neglecting the effect indicated by the above approximations, due to the truncated distribution of $c_r (1 \leq r \leq k - 1)$, we obtain

$$E\{\chi^2\} = 1 + 1 + \cdots + 1 \text{ (averaging first w.r.t. } c_k \text{ for } c_{k-1} \cdots c_1 \text{ fixed and so on)} \tag{10}$$
$$= k$$

χ^4 consists of k terms of the form $[c_r - n_{r-1}(1 - q^{c_{r-1}}(1 - q^{c_{r-1}}))]^4 / [n_{r-1}q^{c_{r-1}}(1 - q^{c_{r-1}})]^2$ and $k(k - 1)$ terms like

$$\frac{[c_r - n_{r-1}(1 - q^{c_{r-1}})]^2}{n_{r-1}q^{c_{r-1}}(1 - q^{c_{r-1}})} \cdot \frac{[c_s - n_{s-1}(1 - q^{c_{s-1}})]^2}{n_{s-1}q^{c_{s-1}}(1 - q^{c_{s-1}})}, \qquad (r \neq s).$$

The contribution to $E(\chi^4)$ from these $k(k - 1)$ terms is $k(k - 1)$, while

$$E\left\{\frac{[c_r - n_{r-1}(1 - q^{c_{r-1}})]^4}{[(n_{r-2} - c_{r-1})(1 - q^{c_{r-1}})q^{c_{r-1}}]^2}\right\}$$

$$= 3 + E^*\left\{\frac{1}{(n_{r-2} - c_{r-1})(1 - q^{c_{r-1}})q^{c_{r-1}}}\right\} - 6E^*\left\{\frac{1}{n_{r-2} - c_{r-1}}\right\}$$

where E^* denotes expectation over $c_1 \cdots c_{r-1}$ $(r > 1)$. So that, in a similar manner to the case of two stages only, provided all $n_{r-1} (1 \leq r \leq k - 1)$ are large enough,

$$\sigma^2(\chi^2) = k(k - 1) + 3k - k^2 + k \text{ smaller terms}$$
$$= 2k + k \text{ small terms.} \tag{11}$$

Hence, for an epidemic lasting k periods, provided the number of susceptibles at each stage, except possibly the last, is large enough, the mean and variance of the χ^2 formed will approximate to those of the true χ^2. Hence, it seems reasonable to suppose that if all the constants were assumed known, the "χ^2" used by Abbey would be at any rate an approximately valid test, which could be used when the initial number of susceptibles, and the number at each succeeding

FIGURE (1a)

$c_0 = 1 \quad n_0 = 100 \quad p = 0.05$

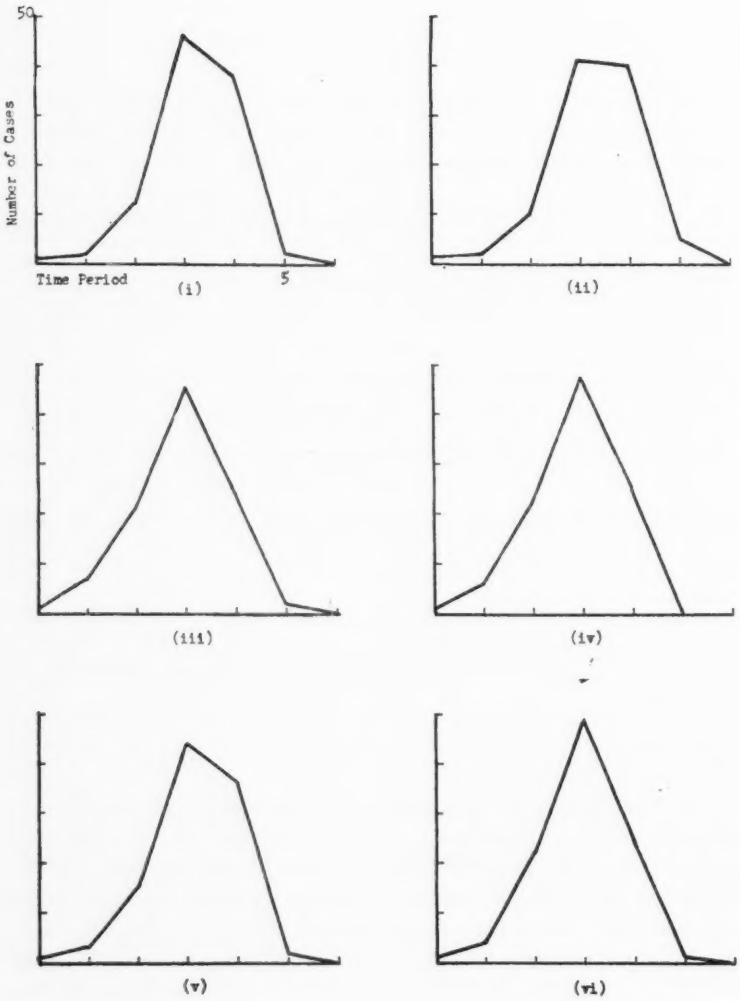
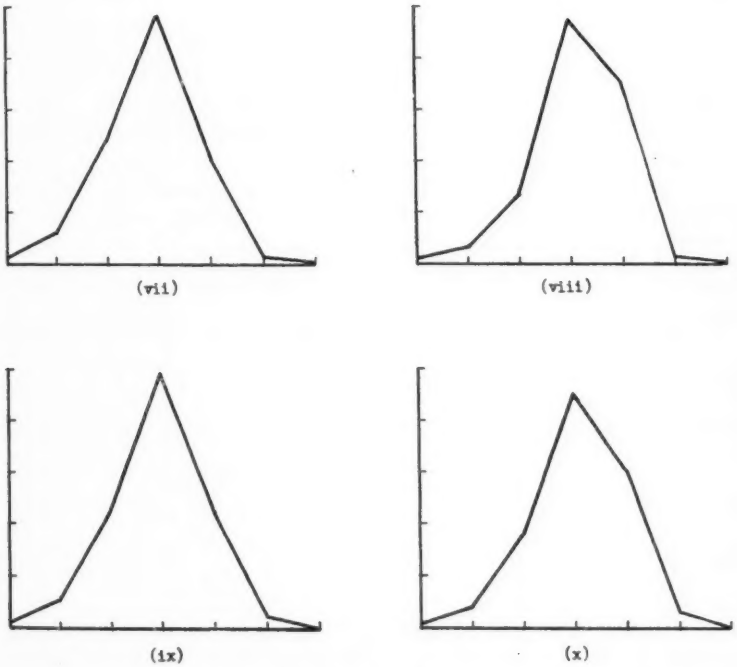


FIGURE (1b)



stage, except possibly the last, is large, and the probability of contact is small, (not much greater than $1/\sqrt{n_0}$).

3. A series of ten mock epidemics were set up in order to investigate the value of χ^2 which would be obtained by testing an "observed" mock epidemic against its theoretical model. All the mock epidemics were based on the assumption that the probability of contact was known and equal to 0.05. The initial population of susceptibles was assumed to be 100, with one initial infectious case.

Taking the probability of infection at time $(t + 1)$ to be $1 - q^{c_t}$ where c_t is the observed number of cases at time t, an epidemic was built up stage by stage using tables of random numbers. The epidemic automatically ended when either the number of cases or number of susceptibles became equal to zero. The epidemics are represented in Figure (1). The expected values at each stage were calculated from the observed values at the previous stage as given in Abbey's paper.

In calculating χ^2, the last stage or stages of the epidemic, where the expected values were too small, were omitted. In practice these were the stages where the expected value was say < 4 as in the usual χ^2 goodness of fit test. As these terms represent the tail end of the epidemic, they are not of great importance in the assessment of the fit of the epidemic to the model, and they may be omitted instead of grouped as in the classical case. None of the values of χ^2 obtained were significantly large. For example the figures for calculation of χ^2 for the first epidemic are given in Table I.

Using equations (7) and (8) the value of the variance of χ^2 for the first two stages was also calculated for $n_0 = 100$; $c_0 = 1$, $p = 0.05$, giving

$$\sigma^2(\chi^2) = 4 + .1602 \tag{12}$$

so that the error term is about 4% of the value of the true variance.

TABLE I
EPIDEMIC 1.

Stage	Observed		Expected	
	c	n	$E(c)$	$E(n)$
0	1	100	1	100
1	2	98	5	95
2	12	96	9.6	88.4
3	46	40	39.6	46.4
4	38	2	36.2	3.8
5	2	0	1.9	0.1

$\chi^2 = 4.47$ (3 d.f.)

4. In addition, there is yet to be considered the effect of estimating a parameter from the data, before applying the χ^2 test.

$$\text{Log likelihood } L = \sum_{r=0}^{k-1} \{\log (n_r)! - \log (c_{r+1})! - \log (n_{r+1})! + c_{r+1} \log (1 - q^{c_r}) + n_{r+1} \log q^{c_r}\} = \sum_{r=1}^{k} L_r .$$

Abbey obtained an estimate of the contact rate from the equation

$$\frac{\partial L}{\partial q} = 0.$$

Considering as before the first two stages of the epidemic only,

$$L = L_1 + L_2 \tag{13}$$

and

$$\frac{\partial L}{\partial q} = \frac{c_0[n_0(1 - q^{c_0}) - c_1]}{q(1 - q^{c_0})} + \frac{c_1[(n_0 - c_1)(1 - q^{c_1}) - c_2]}{q(1 - q^{c_1})} \tag{14}$$

So that the maximum likelihood estimate $\hat{q}$ is obtained from

$$0 = (c_0 n_0 + c_1 n_1)\hat{q}^{(c_0+c_1)} - (c_0 n_0 + c_1 n_1 - c_1 c_2)\hat{q}^{c_0} - (c_0 n_0 + c_1 n_1 - c_0 c_1)\hat{q}^{c_1} + c_0 n_0 + c_1 n_1 - c_0 c_1 - c_1 c_2 .$$

No explicit solution for $\hat{q}$ can be found from this equation.

Following Bartlett (1952) we consider the information function

$$I = I_1 + I_2 \tag{15}$$

$$= E\left\{-\frac{\partial^2 L_1}{\partial q^2}\right\} + E\left\{-\frac{\partial^2 L_2}{\partial q^2}\right\} \tag{16}$$

while, alternatively, the approximate information, obtained by "incomplete" averaging is

$$I' = I_1 + I_2' \tag{17}$$

$$= E\left\{-\frac{\partial^2 L_1}{\partial q^2}\right\} + E_{c_2|c_1}\left\{-\frac{\partial^2 L_2}{\partial q^2}\right\} \tag{18}$$

where $-\partial^2 L_2/\partial q^2$ is averaged over the second stage only, and not over the first.

In the particular example considered

$$-\frac{\partial^2 L}{\partial q^2} = \frac{c_0\{n_0(1 - q^{c_0})^2 - c_1[1 - q^{c_0} - q^{c_0}c_0]\}}{q^2(1 - q^{c_0})^2} + \frac{c_1\{(n_0 - c_1)(1 - q^{c_1})^2 - c_2[1 - q^{c_1} - q^{c_1}c_1]\}}{q^2(1 - q^{c_1})^2} \tag{19}$$

so that

$$I' = \frac{n_0 c_0^2 q^{c_0}}{q^2(1 - q^{c_0})} + \frac{(n_0 - c_1)c_1^2 q^{c_1}}{(1 - q^{c_1})q^2} \tag{20}$$

where c_1 is considered as fixed at the second stage, and

$$I = \frac{n_0 c_0^2 q^{c_0}}{q^2(1 - q^{c_0})} + E_{c_1}\left\{\frac{(n_0 - c_1)c_1^2 q^{c_1}}{(1 - q^{c_1})q^2}\right\}. \tag{21}$$

In the case $c_1 = 0$, the formulae reduce to

$$\frac{\partial L}{\partial q} = \frac{c_0[n_0(1 - q^{c_0}) - c_1]}{q(1 - q^{c_0})}, \qquad I' = \frac{n_0 c_0^2 q^{c_0}}{q^2(1 - q^{c_0})}$$

whereas the completely averaged form I will still of course be

$$I = \frac{n_0 c_0^2 q^{c_0}}{q^2(1 - q^{c_0})} + E_{c_1}\left\{\frac{(n_0 - c_1)c_1^2 q^{c_1}}{(1 - q^{c_1})q^2}\right\}.$$

So that in this case, the use of I' as the information function seems more relevant, for if $c_1 = 0$ there would be no information from any stage later than the first. In the maximum likelihood theory of estimation in large samples, where to the same order of approximation I or $-\partial^2 L/\partial q^2 = -\partial^2 L_1/\partial q^2 - \partial^2 L_2/\partial q^2$ can be used in obtaining a measure of the variance of the estimate, I' is equally good as an approximation, and may be more convenient to use.

If the initial number of susceptibles and the number at the next stage be sufficiently large, considering c_1 as a random variable at the first stage with sample size n_0 , and c_2 as a random variable for the second stage with sample size $n_0 - c_1$ where c_1 *fixed*, the usual maximum likelihood theory in large samples will apply to each stage in turn. The amount of information will be given by I', or to the same order of approximation by $-\partial^2 L/\partial q^2$, or I.

Hence it seems reasonable to assume that estimation of parameters can be carried out as usual, calculating an approximate value of the variance of the estimate from either I, I' or $-\partial^2 L/\partial q^2$.

It may be worth while, and of interest, to check the order of the bias of the estimate. This is done below again considering two stages only, using the formula for the expected value of the likelihood estimate of a parameter (Bartlett, 1953)

$$E\{\hat{\theta}\} = \theta - \left\{\frac{1}{2}E\left\{\frac{\partial^3 L}{\partial \theta^3}\right\} + \frac{\partial I}{\partial \theta}\right\} \Big/ I^2 + \cdots \tag{22}$$

where θ is the parameter to be estimated.

In this formula $E\{\partial^3 L/\partial q^3\}$ can be replaced, within the required order of accuracy, by the unaveraged quantity $\partial^3 L/\partial q^3$ and hence equally well by

$$\text{"}E\text{"}\left\{\frac{\partial^3 L}{\partial q^3}\right\} = E\left\{\frac{\partial^3 L_1}{\partial q^3}\right\} + E_{c_1|c_1}\left\{\frac{\partial^3 L_2}{\partial q^3}\right\}. \tag{23}$$

However

$$\frac{\partial I}{\partial q} = E\left\{\frac{\partial L}{\partial q}\left(-\frac{\partial^2 L}{\partial q^2}\right)\right\} - E\left\{\frac{\partial^3 L}{\partial q^3}\right\}$$

where the right hand side cannot be replaced asymptotically by unaveraged quantities.

An approximate formula for I can be obtained in the form of an infinite series, for

$$I = \frac{n_0 c_0^2 q^{c_0}}{q^2(1-q^{c_0})} + E_{c_1}\left\{\frac{(n_0 - c_1)c_1^2 q^{c_1}}{(1-q^{c_1})q^2}\right\}$$

where

$$\begin{aligned} I_2 &= E_{c_1}\left\{\frac{(n_0 - c_1)c_1^2 q^{c_1}}{q^2(1-q^{c_1})}\right\} \sim \frac{n_0 - \bar{c_1}}{q^2} E_{c_1}\left\{\frac{c_1^2 q^{c_1}}{1-q^{c_1}}\right\} \\ &= \frac{n_0 - \bar{c_1}}{q^2} E_{c_1}\left\{\sum_{r=1}^{\infty} c_1^2 q^{rc_1}\right\} \\ &= \frac{n_0 - \bar{c_1}}{q^2} E_{c_1}\left\{\sum_{r=1}^{\infty} c_1(c_1-1)q^{rc_1} + c_1 q^{rc_1}\right\} \\ &= \frac{n_0^2 q^{c_0}(1-q^{c_0})}{q^2} \sum_{r=1}^{\infty} q^r\{(n_0-1)(1-q^{c_0})q^r[(1-q^{c_0})q^r + q^{c_0}]^{n_0-2} \\ &\qquad + [(1-q^{c_0})q^r + q^{c_0}]^{n_0-1}\} \end{aligned}$$

Hence

$$\begin{aligned} I \sim \frac{n_0 c_0^2 q^{c_0}}{q^2(1-q^{c_0})} + \frac{n_0^2 q^{c_0}(1-q^{c_0})}{q^2} \sum_{r=1}^{\infty} q^r\{n_0(1-q^{c_0})q^r + q^{c_0}\} \\ \cdot[(1-q^{c_0})q^r + q^{c_0}]^{n_0-2}, \end{aligned} \tag{24}$$

whence after differentiating and some algebra

$$\begin{aligned} \frac{\partial I}{\partial q} \sim \frac{n_0 c_0^2 q^{c_0}\{(c_0-2) + 2q^{c_0}\}}{(1-q^{c_0})^2 q^3} + \sum_{r=1}^{\infty} \frac{n_0 q^{c_0} q^r}{q^3}[(1-q^{c_0})q^r + q^{c_0}]^{n_0-3} \\ \cdot[q^{2r} n_0(1-q^{c_0})\{n_0 r(1-q^{c_0}) + (2-c_0) + (3c_0 - 2 - n_0 c_0)q^{c_0}\} \\ + q^r q^{c_0}(1-q^{c_0})\{(3n_0-1)(1-q^{c_0})r + n_0^2 c_0(1-q^{c_0}) \\ + n_0[-3c_0 + 2 - (2-2c_0)q^{c_0}] + 2 + (3c_0-2)q^{c_0}\} \\ + q^{2c_0}\{r(1-q^{c_0}) + n_0(1-q^{c_0})c_0 - 2c_0 + 2 + (3c_0-2)q^{c_0}\}]. \end{aligned} \tag{25}$$

When $c_0 = 1$ these simplify to

$$I = \frac{n_0}{pq} + \frac{n_0^2 p}{q} \sum_{r=1}^{\infty} q^r \{n_0 p q^r + q\}[pq^r + q]^{n_0 - 2}, \tag{26}$$

$$\frac{\partial I}{\partial q} = \frac{n_0(2q-1)}{q^2 p^2}$$

$$+ \sum_{r=1}^{\infty} \frac{n_0^2 q^r}{q^2} [pq^r + q]^{n_0-3}[q^{2r} n_0 p^2 \{n_0 pr + 1 - (n_0 - 1)q\} \tag{27}$$

$$+ q^r pq\{(3n_0 - 1)pr + n_0^2 p - n_0 + 2 + q\} + q^2\{rp + n_0 p + q\}].$$

No summation of the above series has been obtained, but approximate numerical values of I and $\partial I/\partial q$ can be found from these formulae for any particular example.

Consider

$$\sum_{r=1}^{\infty} q^r \{n_0(1 - q^{c_0})q^r + q^{c_0}\}[(1 - q^{c_0})q^r + q^{c_0}]^{n_0 - 2} = \sum_{r=1}^{\infty} f(r)$$

Since $f(x)$ is monotonically decreasing

$$\int_k^{k+1} f(x)\,dx < f(k) < \int_{k-1}^{k} f(x)\,dx$$

hence

$$\int_m^{N+1} f(x)\,dx < \sum_m^N f(r) < \int_{m-1}^{N} f(x)\,dx$$

and as $N \to \infty$

$$\int_m^{\infty} f(x)\,dx < \sum_m^{\infty} f(r) < \int_{m-1}^{\infty} f(x)\,dx;$$

but

$$\int_m^{\infty} q^x[(1 - q^{c_0})q^x + q^{c_0}]^{n_0 - 2}\,dx$$

$$= \frac{1}{\log q} \int_{q^m}^{0} [(1 - q^{c_0})u + q^{c_0}]^{n_0 - 2}\,du \qquad (u = q^x),$$

$$= \frac{1}{(n_0 - 1)\log q(1 - q^{c_0})} \{(q^{c_0})^{n_0 - 1} - [(1 - q^{c_0})q^m + q^{c_0}]^{n_0 - 1}\}.$$

Similarly

$$\int_m^\infty q^{2x}[(1-q^{c_0})q^x + q^{c_0}]^{n_0-2}\,dx$$

$$= \frac{1}{\log q}\int_{q^m}^0 u[(1-q^{c_0})u + q^{c_0}]^{n_0-2}\,du$$

$$= \frac{1}{\log q}\left\{-\frac{q^m[(1-q^{c_0})q^m + q^{c_0}]^{n_0-1}}{(n_0-1)(1-q^{c_0})}\right.$$

$$\left. - \frac{1}{(n_0-1)(1-q^{c_0})}\int_{q^m}^0 [(1-q^{c_0})u + q^{c_0}]^{n_0-1}\,du\right\}$$

$$= \frac{-q^m[(1-q^{c_0})q^m + q^{c_0}]^{n_0-1}}{(n_0-1)(1-q^{c_0})\log q} - \frac{\{(q^{c_0})^{n_0} - [(1-q^{c_0})q^m + q^{c_0}]^{n_0}\}}{n_0(n_0-1)(1-q^{c_0})^2\log q}.$$

Hence

$$\int_m^\infty q^x\{n_0(1-q^{c_0})q^x + q^{c_0}\}[(1-q^{c_0})q^x + q^{c_0}]^{n_0-2}\,dx$$

$$= \frac{-[(1-q^{c_0})q^m + q^{c_0}]^{n_0-1}q^m}{\log q}$$

so

$$\frac{-[(1-q^{c_0})q^m + q^{c_0}]^{n_0-1}q^m}{\log q} < \sum_m^\infty f(r) < \frac{-[(1-q^{c_0})q^{m-1} + q^{c_0}]^{n_0-1}q^{m-1}}{\log q}$$

and so

$$\frac{n_0^2q^{c_0}(1-q^{c_0})}{q^2}\sum_{r=1}^{m-1} f(r) - \frac{n_0^2q^{c_0}(1-q^{c_0})[(1-q^{c_0})q^m + q^{c_0}]^{n_0-1}q^m}{q^2\log q} < I_2$$

$$< \frac{n_0^2q^{c_0}(1-q^{c_0})}{q^2}\sum_1^{m-1} f(r) \qquad (28)$$

$$- \frac{n_0^2q^{c_0}(1-q^{c_0})[(1-q^{c_0})q^{m-1} + q^{c_0}]^{n_0-1}q^{m-1}}{q^2\log q}$$

or, for $c_0 = 1$,

$$\frac{n_0^2p}{q}\sum_1^{m-1} f(r) - \frac{n_0^2p[pq^m + q]^{n_0-1}q^m}{q\log q} < I_2 < \frac{n_0^2p}{q}\sum_1^{m-1} f(r) \qquad (29)$$

$$- \frac{n_0^2p[pq^{m-1} + q]^{n_0-1}q^{m-1}}{q\log q}.$$

Even for quite small values of m this should give a satisfactory approximation for I_2.

For example, when $n_0 = 100$, $c_0 = 1$, $p = 0.05$,

$$\text{for } m = 5 \qquad 8470 < I_2^{(5)} < 9130$$

$$\text{for } m = 10 \qquad 8700 < I_2^{(10)} < 8890.$$

So that a reasonably good approximation for I_2 is $\frac{1}{2}(8700 + 8890) = 8795$. An approximation for the infinite series for $\partial I_2/\partial q$ cannot be found in this way. However an approximate numerical value can be found using extrapolation methods. For example, writing

$$\frac{\partial I_2}{\partial q} = \frac{n_0^2}{q^2}\sum_{r=1}^{\infty} u(r) = \frac{n_0^2}{q^2}F(\infty) \qquad \text{say, where} \qquad \sum_{r=1}^{m} u(r) = F(m).$$

By direct computation ($n_0 = 100$, $c_0 = 1$, $p = 0.05$) $F(3) = 11.31$: $F(4) = 15.05$: $F(6) = 21.27$: $F(12) = 31.37$ so that if $f(12/r) = F(r)$ $f(1) = 31.37$: $f(2) = 21.27$: $f(3) = 15.05$: $f(4) = 11.31$. Using Gregory's formula for interpolation and extrapolation ($\theta < 0$) (see for example Jeffreys and Jeffreys "Mathematical Physics" p. 242)

$$y(x_0 + \theta h) = y_0 + \theta\Delta y_0 + \frac{\theta(\theta - 1)}{2!}\Delta^2 y_0 + \cdots$$

where Δy_0, $\Delta^2 y_0$ etc. are the differences $\Delta y_0 = y_1 - y_0$, $\Delta^2 y_0 = \Delta y_1 - \Delta y_0$ etc. The difference table reads

n	$f(n)$	$\Delta f(n)$	$\Delta^2 f(n)$	$\Delta^3 f(n)$
1	31.37			
		−10.10		
2	21.27		3.88	
		− 6.22		−1.40
3	15.05		2.48	
		− 3.74		
4	11.31			

So that, at least as far as the above figures indicate,

$$f(0) = 31.37 + 10.10 + 3.88 + 1.40$$
$$= 46.75 \sim 47.$$

So

$$F(\infty) \sim 47 \qquad \text{and} \qquad \frac{\partial I_2}{\partial q} \sim 520{,}800.$$

I_1 , $\partial I_1/\partial q$ can be obtained by direct calculation, and in this example

$$I_1 = \frac{n_0 c_0^2 q^{c_0}}{q^2(1 - q^{c_0})} = 2{,}105, \qquad \frac{\partial I_1}{\partial q} = \frac{n_0 c_0^2 q^{c_0}\{(c_0 - 2) + 2q^{c_0}\}}{(1 - q^{c_0})q^3}$$

$$= 39{,}890.$$

To calculate $E\{\partial^3 L/\partial q^3\}$,

$$\frac{\partial^3 L_1}{\partial q^3} = \frac{c_0}{q^3(1 - q^{c_0})^3}\{2n_0(1 - q^{c_0})^3 - c_1[2 + q^{c_0}(c_0 + 1)(c_0 - 4) + q^{2c_0}(c_0 + 1)(c_0 + 2)]\}, \tag{30}$$

$$E_{c_1}\left\{\frac{\partial^3 L_1}{\partial q^3}\right\} = \frac{-c_0^2 n_0 q^{c_0}}{q^3(1 - q^{c_0})^2}\{(c_0 - 3) + q^{c_0}(c_0 + 3)\},$$

and similarly for $\partial^3 L_2/\partial q^3$. So

$$\text{“}E\text{”}\left\{\frac{\partial^3 L}{\partial q^3}\right\} = E\left\{\frac{\partial^3 L_1}{\partial q^3}\right\} + E_{c_2|c_1}\left\{\frac{\partial^3 L_2}{\partial q^3}\right\}$$

$$= \frac{-c_0^2 n_0 q^{c_0}}{q^3(1 - q^{c_0})^2}\{(c_0 - 3) + q^{c_0}(c_0 + 3)\} \tag{31}$$

$$- \frac{c_1^2(n_0 - c_1)q^{c_1}}{q^3(1 - q^{c_1})^2}\{(c_1 - 3) + q^{c_1}(c_1 + 3)\}$$

When $n_0 = 100$, $c_0 = 1$, $p = 0.05$ and $c_1 = 3$ say

$$\text{“}E\text{”}\left\{\frac{\partial^3 L}{\partial q^3}\right\} = -79{,}778.4 - 223{,}031.7$$

$$\sim -302{,}810.$$

Hence

$$\text{Bias} \sim \frac{-\left\{\frac{1}{2}\,\text{“}E\text{”}\left\{\frac{\partial^3 L}{\partial q^3}\right\} + \frac{\partial I}{\partial q}\right\}}{I^2} \qquad (I = I_1 + I_2) \tag{32}$$

$$\sim -0.0034.$$

Alternatively using the approximate information function I' in the denominator,

$$I' = I_1 + I_2' = \frac{n_0 c_0^2 q^{c_0}}{q^2(1 - q^{c_0})} + \frac{(n_0 - c_1)c_1^2 q^{c_1}}{(1 - q^{c_1})q^2} \qquad \text{(equation (20))}$$

$$= 2{,}105 + 5{,}815.$$

$$\text{Bias} \sim \frac{-\left\{\frac{1}{2}\,\text{“}E\text{”}\left\{\frac{\partial^3 L}{\partial q^3}\right\} + \frac{\partial I}{\partial q}\right\}}{I'^2} \sim -0.0065. \tag{33}$$

Both approximations indicate that the bias is sufficiently small, of order $1/n_0$ as expected.

As far as the χ^2 is concerned, the additional effect of the estimation will then be as usual, to subtract from the original χ^2 a quantity with a distribution approximately that of a χ^2 with one degree of freedom. Hence the χ^2 test can still be used, when a parameter has been estimated, provided that as usual the number of degrees of freedom is reduced by one.

The extension to more than two stages follows fairly obviously. The maximum likelihood theory in large samples will still apply to each stage in turn, provided the number of susceptibles at the previous stage is large enough, and the number of cases c_r is considered as a random variable at stage r, and fixed at stage $r + 1$.

Hence, provided the number of susceptibles at each stage except possibly the last is large enough, the χ^2 used by Abbey will still be an approximately valid test of goodness of fit, when the contact rate is estimated from the data. The estimation of the further parameter n_0 as carried out later in Abbey's paper, should introduce no further difficulty. The use of c_r first as a random variable then as a known fixed quantity appears at first sight rather inconsistent, but there seems no reason why c_r should not be considered in this way in the present case.

I should like to thank Professor Bartlett for his advice and encouragement during the preparation of this paper. I should also like to thank Mrs. Linnert for her assistance in computing the mock epidemics.

SUMMARY

A brief description of how Dr. Helen Abbey in "An Examination of the Reed-Frost Theory of Epidemics" uses χ^2 to test epidemics against mathematical models is given, and the difference between this test and the usual χ^2 Goodness of Fit Test pointed out. The distribution of the χ^2, assuming all constants known, is investigated by evaluating the approximate value of its mean and variance, when the χ^2 test is applied to the first two stages of the epidemic only. The mean and variance are shown to be approximately equal to those of the usual χ^2. The extension to more than two stages is also indicated. The value of the variance of the χ^2 determined from the first two stages is then calculated numerically for a mock epidemic.

The application of the maximum likelihood theory of estimation in large samples, to the first two stages of the epidemic is considered, including in particular the effect of estimation of the contact rate q

on χ^2. The approximate bias of the maximum likelihood estimate of q is evaluated numerically for a mock epidemic. Finally the extension of the theory of estimation to more than two stages is indicated.

REFERENCES

Abbey, H. "An examination of the Reed-Frost theory of Epidemics", *Human Biology*, 1952, 3, 201–233.

Bartlett, M. S. "The statistical significance of odd bits of information", *Biometrika*, 1952, 39, 228–237.

"Approximate confidence intervals II. More than one unknown parameter", *Biometrika*, 1953, 40, 306–317.

Jeffreys, H. and B. S. "Mathematical Physics", Cambridge University Press. 1946.

MISCLASSIFICATION IN 2 × 2 TABLES

IRWIN BROSS

Cornell University Medical College

INTRODUCTION

Although in the derivation of the chi-square test (and related techniques) the classifications in the 2 × 2 table are assumed to be correct, there are many practical problems where mistakes in classification are going to be made. The important question then arises: What effects will the misclassification have on conclusions drawn from the usual significance tests? The principal purpose of this paper is to answer this question.

In the medical field there are some classifications that involve almost no risk of error, for example, the categories "lived" and "died". On the other hand, in more complex diagnoses the clinician realizes that there is a considerable risk of error, a risk that may vary a great deal depending on the disease under study, the existence and availability of diagnostic tests, and other factors. For some diseases the principal misclassification will be in the direction of missing some of the actual cases. In other diseases there may be a risk of misdiagnosis in the other direction due to the existence of diseases which "mimic" the characteristics of the disease under study.

Even when more precise classification is possible there may be time or cost factors that necessitate the use of cruder classification methods. In a large survey of mental health, for example, it will rarely be possible to have actual psychiatric diagnoses on all respondents. Instead it will be necessary to classify respondents as "Health Problems" or "Not Health Problems" on the basis of information in a question schedule. One possible procedure in such situations is to construct a mental health scale (using a subsample of "problem" and "non-problem" cases) and thereafter classify individuals by whether they fall above or below some *critical point* on the mental health scale. Such a procedure may well lead to appreciable misclassifications in both directions.

Research workers have quite different opinions as to what can be done in the analysis of data which are subject to misclassifications. Some feel that in a large series the effects will "cancel out" and therefore it is proper to use the usual analysis. At the other extreme are the research workers who argue that the data are biased and therefore no conclusions can be drawn. The findings of this paper represent an intermediate position.

In discussing the effects of misclassification it is important to remember that a given body of data may be suitable for one purpose but not for another. In the mental health survey, for instance, we may want to see if there are different proportions of "problem" cases in two sub-populations (such as ethnic or economic groups) and for this purpose *significance tests* would be used. We may also want to estimate the proportion of "problem" cases in a specified sub-population and for this purpose *confidence intervals* might be used.

Misclassification may present a more serious problem in the case of estimates than in the case of significance tests.

A second point to note is that the performance of a statistical method may be judged by *several* standards. One standard is that a significance test should be *valid*. In other words the risk of a false assertion that "two populations are different" (Type I error) should be some preassigned value such as 5%.

If the same classification system is used in the two populations then misclassification will not affect the validity of the significance test.

Thus, if the experimenter performs chi-square tests in the usual fashion the risk of making false Type I assertions will not be increased by the misclassification. However, there is a price that must be paid for misclassification.

Misclassification tends to reduce the power of the test.

A second standard for judging a significance test is that it should be *powerful*. In other words, it should be able to detect real differences between two populations. Practically speaking the reduction of power corresponds to reduction in effective sample size. Thus in a survey with 2,000 respondents, the misclassification might be equivalent to the "loss" of 500 interviews.

Naturally the extent of loss will depend on the amount of misclassification and a table will be presented which will indicate the magnitudes of the losses in different situations. It is also possible to reduce the loss by taking account of misclassification when the study is planned, for example, by the choice of the *critical point* when a scale is used to make the classifications.

ANALYTICAL MODEL

In order to justify the statements in the introduction and to obtain a more quantitative picture of the effects of misclassification, an analytical model will be set up. As a first step, we will consider effects of misclassification on a sample from a *single* population.

Suppose that a sample of n individuals is drawn from the population. Let u be the *actual* number of "problem" individuals in the sample. Let v be the number of "problem" individuals who are *incorrectly* classified as "non-problem" cases. Let w be the number of "non-problem" individuals who are *incorrectly* classified as "problem" cases. Finally let x be the *apparent* number of "problem" individuals (i.e., the number according to the inaccurate classification system). It follows that $x = u - v + w$.

Now let p be the proportion of "problem" cases in the *population*. Let θ be the probability of misclassifying a "problem" individual. Let ϕ be the probability of misclassifying a "non-problem" individual. Parame'ers θ and ϕ are characteristics of the *classification system*.

As a result of misclassification the estimated proportion of "problem" cases will not in general be p as before. Instead the expected value of x/n will be:

$$E\left(\frac{x}{n}\right) = p - p\theta + q\phi \qquad \text{where} \qquad q = (1 - p)$$

Although the expectation has been changed, x is a binomial variable and the variance of x obtained* in the usual way, will be:

$$V(x) = n(p - p\theta + q\phi)(q + pe - q\phi)$$

If now:

$$K = 1 - \theta + \frac{q}{p}\phi \tag{1.01}$$

$$K' = (1 - \phi - \theta)^2 + \frac{\theta(1 - \theta)}{q} + \frac{\phi(1 - \phi)}{p} \tag{1.02}$$

then if n is large, the number of "problem" cases, x, will be approximately normally distributed with mean

$$E(x) = npK$$

and variance

$$V(x) = npqK'.$$

If there is no misclassification, $K = K' = 1$.

*This simplified derivation was suggested by Referee I.

If, as will usually be the case, θ and ϕ are small fractions whose squares will be negligible, K' can be approximated by K'' where:

$$(1.03) \qquad K'' = 1 - 2\phi - 2\theta + \frac{\theta}{q} + \frac{\phi}{p} = 1 + \theta\left(\frac{1}{q} - 2\right) + \phi\left(\frac{1}{p} - 2\right)$$

ESTIMATION

An examination of K and K' tells the story concerning the effects of misclassification on estimates. *In the following discussion it will be assumed that p is less than $\frac{1}{2}$.* If p is greater than $\frac{1}{2}$ the roles of the two types of misclassification (i.e., of θ and ϕ) will be interchanged. *Moreover it will also be assumed that θ and ϕ are of the same order of magnitude*—this being the usual situation in practice. If θ is much larger than ϕ, the remarks may not apply.

If x/n is used as an estimate of p, a bias may be introduced by the misclassification. The bias is:

$$(1.04) \qquad B = q\phi - p\theta = (K - 1)p$$

so that K is a measure of the relative bias. Under the above assumptions, the chief bias will arise from misclassifications of the "non-problem" cases and the bias will get progressively worse as p becomes smaller and smaller. For example, if there is a 5% risk of misclassification in either direction ($\theta = \phi = .05$) and if there are 10% "problem" cases in the population ($p = .10$), then by (1.01):

$$K = 1 - .05 + \frac{.90}{.10}(.05) = 1.40$$

so that the proportion of problem cases will be overestimated by 40%. If estimates of θ and ϕ are available from the data used to construct the classification system, the estimates might be adjusted by dividing by K.

An alternative procedure would be to choose the critical point on the classification scale so that:

$$(1.05) \qquad \phi/\theta = p/q.$$

With this choice of the critical point unbiased estimates would be obtained. Unfortunately the value of p would not be known in advance so in practice (1.05) could only be approximated. Moreover, if (1.05) is used for the whole population the estimates for sub-populations might still be biased.

The misclassification not only introduces a bias but also affects the variance of the estimate. When an estimate is biased it is necessary to

distinguish clearly between the variance (which is the second moment about the expected value) and the mean square error (which is the second moment about the true value). With biased estimates the mean square error is a better index of efficiency.

The quantity K' measures the effect of misclassification on the variance of the estimate, however, it is easier to see the effects of misclassification if the approximate quantity K'' (1.03) is examined. It can be seen that as p becomes smaller the variance tends to increase. Note that the effect of misclassification of the "problem" cases (i.e. θ) is to *reduce* the variance. As p becomes smaller and smaller this effect becomes unimportant. On the other hand, the effect of misclassifications of the "non-problem" cases (i.e. ϕ) is to increase the variance and this effect increases as p becomes smaller and smaller. The accepted measure of efficiency is the reciprocal of $K' + D$ where

$$D = \frac{(q\phi - p\theta)^2 n}{pq}$$

It will be noted that the efficiency so defined will *decrease* as the sample size increases. Thus in a sample of 100 with $\theta = \phi = .05$ and $p = .10$, the value of K' is 1.34 and the value of D is 1.78. Therefore the index of efficiency, the reciprocal of $1.34 + 1.78 = 3.12$, will be 32%. A 32% efficiency has the interpretation that about $\frac{2}{3}$ of the original 100 observations are "lost" due to the misclassification.

The use of (1.05) to fix the critical point will not only help to unbias the estimates, but it will also greatly improve the efficiency. However, in many studies the critical point will necessarily be determined by other considerations than (1.05) and an appreciable bias may be introduced. As the foregoing examples indicate, relatively small misclassification rates can introduce very serious biases and lead to very low efficiencies. While in theory an adjustment can be made either by calculating K or by using (1.04) to calculate B, in practice the quantities θ and ϕ are required and only crude estimates of these quantities may be available. The research worker then has an unpleasant choice between an unadjusted estimate and a rather arbitrary adjustment. In this situation, it may be best to present both estimates together with a clear statement of the difficulty. The "standard" confidence intervals should *not* be used here.

SIGNIFICANCE TESTS

The principal interest of this paper lies in the comparison of samples from two populations (denoted by subscripts 1 and 2). To compare

the proportion of problem cases, the statistic

$$d = \frac{x_1}{n_1} - \frac{x_2}{n_2}$$

would be used.

An especially important practical situation is the one where the same classification system is used in both samples. It will be shown that for this situation, the significance test remains valid even though the misclassification is altogether ignored. The comfort that the research worker can draw from this result is somewhat mitigated by the fact that the significance test, though valid, becomes less powerful.

In the previous section it has been shown that the expected value of x when misclassification occurs is Kp and the variance is $nKp\ (1 - Kp)$ where $K = 1 - \theta - (q/p)\theta$. In terms of K' (see equation (1.02)) the variance of x is $npqK'$ so:

$$pqK' = Kp(1 - Kp)$$

If now, x_1 and x_2 are independent samples from two populations where the same classification system has been used, then

$$E(d) = \frac{n_1 p_1 K}{n_1} - \frac{n_2 p_2 K}{n_2} = K(p_1 - p_2)$$

and under the null hypothesis (i.e. $p_1 = p_2$) the expectation of d is 0 whether there is misclassification or not.

The variance of d under the null hypothesis will be:

$$V(d) = \frac{pqK'}{n_1} + \frac{pqK'}{n_2} = pqK'A$$

where

$$A = \frac{1}{n_1} + \frac{1}{n_2}.$$

If the misclassification is altogether ignored, the variance would be estimated by first estimating p by

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

and then estimating pq/N by $\hat{p}\hat{q}/(N - 1)$ where $N = n_1 + n_2$ so that

$$\hat{V}(d) = \hat{p}\hat{q}\,\frac{N}{N - 1}\,A.$$

While it is known that the above is an unbiased estimate of the variance where there is no misclassification, the question that now must be answered is: What is the effect of misclassification on this estimate?

When misclassification occurs

$$E\hat{p}\hat{q} = E\hat{p}(1 - \hat{p}) = E\hat{p} - E\hat{p}^2 = Kp - E\hat{p}^2$$

but

$$V\hat{p} = \frac{pqK'}{N} = E\hat{p}^2 - (E\hat{p})^2$$

so that

$$E\hat{p}^2 = (E\hat{p})^2 + \frac{pqK'}{N} = K^2p^2 + \frac{pqK'}{N}$$

Hence

$$E\hat{p}\hat{q} = Kp - K^2p^2 - \frac{pqK'}{N}$$

and

$$E\hat{V} = \frac{N}{N-1} A\left\{Kp(1 - Kp) - \frac{pqK'}{N}\right\}$$

while the actual variance is

$$V = pqK'A.$$

Since $Kp(1 - Kp) = pqK'$

$$E\hat{V} = \frac{N}{N-1} A\left\{pqK'\left(1 - \frac{1}{N}\right)\right\} = pqK'A.$$

so that even when misclasification is present, an unbiased estimate of the variance is obtained. This means that the existence of misclassification does not affect the validity of a significance test based on d, and since the ordinary Chi-square test is equivalent to a test based on d, it follows that the validity of the Chi-square test is also not affected by ignoring misclassification.*

We do not get off "scot-free" however. Although the tests are valid the power may be drastically reduced. In calculating the power of the test *effective* sample sizes of $1/K'$ times the actual sample sizes should be used. Table I presents values of the efficiency $(1/K')$ for representative values of θ, ϕ, and p.

*I am indebted to the referees for improvements in this proof.

TABLE I

THE INFLUENCE OF MISCLASSIFICATION ON EFFICIENCY

θ	p	ϕ .01	ϕ .02	ϕ .05	ϕ .10	ϕ .20
.01	.05	.856	.744	.542	.384	.261
	.10	.934	.871	.730	.587	.447
	.15	.963	.923	.825	.712	.587
	.20	.978	.951	.882	.797	.696
.05	.05	.884	.765	.552	.389	.262
	.10	.966	.898	.748	.597	.451
	.15	.995	.951	.846	.725	.593
	.20	1.008	.978	.903	.812	.703
.10	.05	.922	.793	.566	.395	.264
	.10	1.009	.934	.771	.610	.457
	.15	1.037	.989	.873	.743	.601
	.20	1.048	1.015	.932	.832	.713

A study of this table will give the research worker a better "feeling" for the effects of misclassification. It shows, for example, that when the "problem" case is a rare event then even small rates of ϕ (misclassification of "non-problem" cases) can result in the loss of about half of the data. Insofar as the significance test is concerned, the effect of misclassification of "problem" cases is not very great but it should be remembered that estimates will be much more seriously affected by this kind of misclassification.

If the test turns out to be significant we may wish to use d to estimate the difference in population proportions. This estimate will be biased by the factor K, however, and the previous remarks concerning estimation will apply here as well.

SMALL SAMPLES

Since we have been primarily interested in the situation where θ, ϕ, and p are small, it may occur to some readers that the normal distributions hitherto assumed might not be very good approximations in samples of a realistic size. Therefore it seems worthwhile to examine what happens in small samples in a simple special case.

Since the quantity v is going to be a fraction of u, it is relatively unimportant and for the sake of simplicity we will now consider only the quantities u and w, so:

$$x = u + w$$

As ϕ and p are small, it is appropriate to regard u and w as following Poisson distributions with expectations np and $(n - u)\phi$ respectively. As a further approximation we note that u is small relative to n, so that we shall take the expectation of w as simply $n\phi$. Under these assumptions the joint distribution of u_1 , u_2 , and w_1 , w_2 will be the product of four Poisson distributions. Thus the distribution of u_1 would be:

$$P(u_1) = \frac{e^{-n_1 p_1}(n_1 p_1)^{u_1}}{u_1!}$$

The joint distribution, $P(u_1 , u_2 , w_1 , w_2)$ can be rewritten in another form which will be more useful. Let $t = x_1 + x_2$. Then:

$$P(u_1 , u_2 , w_1 , w_2) = P(t)P(x_1 , x_2 \mid t)P(u_1 , w_1 \mid x_1)P(u_2 , w_2 \mid x_2)$$

where:

$$P(t) = \frac{e^{-A}A^t}{t!} \qquad A = n_1 p_1 + n_1\phi + n_2 p_2 + n_2\phi$$

$$P(x_1 , x_2 \mid t) = \frac{t!}{x_1!x_2!}P^{x_1}Q^{x_2} \qquad P = \frac{n_1(p_1 + \phi)}{n_1(p_1 + \phi) + n_2(p_2 + \phi)}$$

$$P(u_i , w_i \mid x_i) = \frac{x_i!}{u_i!w_i!}P_i^{u_i}Q_i^{w_i} \qquad P = \frac{n_i p_i}{n_i p_i + n_i\phi} \qquad i = 1, 2$$

The important point to note is that if there is no difference between populations the conditional distribution, $P(x_1 , x_2 \mid t)$ is a binomial distribution which does not depend on either p or ϕ, since:

$$P = \frac{n_1}{n_1 + n_2}$$

If there is no misclassification ($\phi = 0$), the same distribution of x_1 and x_2 is obtained. Consequently the validity of the significance test based on $P(x_1 , x_2 \mid t)$ is undisturbed by misclassification. This result agrees with the large sample result obtained previously.

A 2 × 2 FACTORIAL WITH PAIRED COMPARISONS[1]

ROBERT M. ABELSON[2] AND RALPH ALLAN BRADLEY

Virginia Agricultural Experiment Station of the Virginia Polytechnic Institute

1. INTRODUCTION

Bradley and Terry [2] have presented a method of paired comparisons depending on ranking within the incomplete blocks of size two. The method may be regarded as a generalization of the sign test. The model has recently been reviewed [3, 9] and its appropriateness examined with reference to certain experimental results. It is the purpose of this paper to consider treatments that form a factorial set in paired comparisons.

A brief review of the basic model for paired comparisons will assist in understanding the modifications to be introduced. t parameters are defined for t treatments, $T_1, \cdots, T_t$, in a paired comparisons experiment with n repetitions of the $\binom{t}{2}$ comparisons. The treatment parameters are $\pi_1, \cdots, \pi_t$ with $\pi_i \geq 0$, $i = 1, \cdots, t$, and $\Sigma\pi_i = 1$. It is postulated that the parameters are such that, if X_i is an observation on T_i, then

$$P(X_i > X_j) = \pi_i/(\pi_i + \pi_j) \tag{1}$$

in the comparison of T_i with T_j. The rank of T_i when compared with T_j in the kth repetition is r_{ijk} and Σr_i is used as an abbreviation for the total sum of ranks (over all comparisons and all repetitions) for T_i.

The general likelihood function is

$$L(\pi_i) = \prod_{i=1}^{t} \pi_i^{2n(t-1)-\Sigma r_i} \prod_{i<j} (\pi_i + \pi_j)^{-n}. \tag{2}$$

To test the null hypothesis of treatment equality,

$$H_0 : \pi_i = 1/t, \qquad i = 1, \cdots, t,$$

[1] A condensation of a thesis presented by the first author to the Virginia Polytechnic Institute in partial fulfillment of the requirements for the M.S. Degree in Statistics. Prepared under the direction of the second author and in part under a Research and Marketing Act Contract, Project RM: e-629-1, with the Bureau of Agricultural Economics.

[2] Now with American Machine and Foundry Company, Electronics Division, Boston, Massachusetts.

against the general alternative,

$$H_a : \pi_i \neq \frac{1}{t} \qquad \text{for some values of } i,$$

the test statistic used is[3]

$$(3) \qquad B_1 = \sum_{i<j} n \log (p_i + p_j) - \sum_i \{2n(t-1) - \Sigma r_i\} \log p_i$$

where p_i is the maximum likelihood estimator of π_i using (2). Some tables of the distribution of B_1 for small values of t and of n have been published [2] and extended tables have been submitted for publication [5]. Values of $p_1, \cdots, p_t$ are included in the tables and they were obtained using the equations,

$$(4) \qquad \frac{\{2n(t-1) - \Sigma r_i\}}{p_i} - n \sum_{j \neq i} \frac{1}{p_i + p_j} = 0, \qquad i = 1, \cdots, t,$$

and

$$(5) \qquad \sum_i p_i = 1.$$

An approximate test is available since

$$(6) \qquad -2 \ln \lambda_1 = nt(t-1) \ln 2 - 2B_1 \ln 10$$

has the distribution of χ^2 with $(t-1)$ degrees of freedom.

Thurstone [11], in discussing paired comparisons, postulated an additive subjective continuum on which treatments have location points S_i. The variates are taken to be normally distributed about their location points with a common variance. It follows from his work[4] that

$$(7) \qquad P(X_i > X_j) = \frac{1}{\sqrt{2\pi}} \int_{-(S_i - S_j)}^{\infty} e^{-y^2/2}\, dy.$$

In the model reviewed above, it is shown [4] that

$$(8) \qquad P(X_i > X_j) = \frac{1}{4} \int_{-(\ln \pi_i - \ln \pi_j)}^{\infty} \operatorname{sech}^2 \frac{y}{2}\, dy$$

and $\ln \pi_i$ may be taken as a location parameter for T_i. We shall use this result in formulating the model for factorial treatments.

The model for paired comparisons is generalized by supposing that the experiment is conducted in g groups of repetitions with n_u repetitions in the uth group. It may not be realistic to assume that the treatment

[3] We use log and ln to represent logarithms to bases 10 and e respectively.

[4] This summary of Thurstone's model is most clearly presented by Mosteller [10].

parameters are identical for each group. Then it is postulated that parameters $\pi_{1u}, \cdots, \pi_{tu}$ exist for the uth group, $u = 1, \cdots, g$. Two additional tests are given for this situation. If we consider the null hypothesis,

$$H_0 : \pi_{iu} = \frac{1}{t} \qquad \text{for all } i \text{ and } u$$

and the alternative hypothesis,

$$H_a : \pi_{iu} \neq \frac{1}{t} \qquad \text{for some } i \text{ and } u,$$

the test statistic is

$$B_1^c = \sum_u B_1^u \tag{9}$$

where B_1^u corresponds to B_1 in (3) for the uth group. For an approximate test,

$$-2 \ln \lambda_1^c = -2 \sum_u \ln \lambda_1^u \tag{10}$$

has the χ^2-distribution with $g(t-1)$ degrees of freedom. λ_1^u is related to B_1^u through (6).

A test of agreement or of no interaction of treatments with groups may be obtained. The null hypothesis is

$$H_0 : \pi_{iu} = \pi_i \qquad \text{for all } i \text{ and } u$$

with the alternative

$$H_a : \pi_{iu} \neq \pi_i \qquad \text{for some } i \text{ and } u.$$

The statistic,

$$2(B_1 - B_1^c) \ln 10 \tag{11}$$

approximates χ^2 with $(g-1)(t-1)$ degrees of freedom. B_1 is computed using (3) with $n = \sum_u n_u$ repetitions under the assumptions of the null hypothesis.

2. THE MODEL FOR FACTORIALS

Let us now consider the case where we wish to study several different factors in a single experiment. We shall designate the factors by $A_1, \cdots, A_m$ and their numbers of levels by $a_1, \cdots, a_m$ respectively. With the treatments in factorial array, there are $t = \pi_{s=1}^m a_s$ treatment combinations. To illustrate, we might think of a taste testing experi-

ment for preferences using several amounts of a number of condiments.

For the general factorial, we assume that the levels of each factor have true ratings whose values are typified by $\alpha_{1s}, \cdots, \alpha_{a_s s}$ for the sth factor. This simply means that α_{vs} is a parameter representing the effect of the vth level of the factor A_s. It is further assumed that $\alpha_{vs} \geq 0$ for all v and s and that

$$\sum_{v=1}^{a_s} \alpha_{vs} = 1, \qquad s = 1, \cdots, m. \tag{12}$$

The condition (12) is added to provide a determinate solution for equations yielding estimates of these parameters and it is analogous to the condition usually imposed in the analysis of variance that requires that the sum of the effects due to the levels of a factor in the additive model be zero.

We shall use $T(\alpha_v)$ to represent the treatment composed of the m factors, $A_1, \cdots, A_m$, with levels in the vector, $(\alpha_v) = (\alpha_{v_1 1}, \cdots, \alpha_{v_m m})$. That is, $T(\alpha_v)$ is a treatment combination with the factor A_1 at level v_1, factor A_2 at level v_2, and so on. In the model for paired comparisons reviewed above, the rating or parameter for $T(\alpha_v)$ may be denoted by $\pi(\alpha_v)$ and, for the factorial,

$$\pi(\alpha_v) = \prod_{i=1}^{m} \alpha_{v_i i}. \tag{13}$$

It will be seen that π_i, the parameter associated with treatment T_i in the general model for paired comparisons, is now replaced by the parameter, $\pi(\alpha_v)$, for treatment $T(\alpha_v)$. The composition of $\pi(\alpha_v)$ is then given by (13).

We define $r(\alpha_v, \alpha_w)_k$ to be the rank of $T(\alpha_v)$ when compared with some other treatment, $T(\alpha_W)$, in the kth repetition of the experiment. Thus, $r(\alpha_v, \alpha_w)_k$ replaces r_{ijk} in the general model and it is also consistent to next define $\Sigma r(\alpha_v)$ to be the total sum of ranks of $T(\alpha_v)$ over all comparisons involving $T(\alpha_v)$ in the k repetitions of the experiment. With these definitions, substitution in the likelihood function (2) yields

$$L\{\pi(\alpha_v)\} = \prod_{(\alpha_v)} \pi(\alpha_v)^{2n(t-1)-\Sigma r(\alpha_v)} \prod_{(\alpha_v)\neq(\alpha_w)} \{\pi(\alpha_v) + \pi(\alpha_w)\}^{-n/2}. \tag{14}$$

$\prod_{(\alpha_v)}$ denotes a product over all distinct vectors, (α_v), or for all parameters, $\pi(\alpha_v)$, associated with the t treatments $T(\alpha_v)$ where t is defined in terms of the numbers of levels of the factors. $\prod_{(\alpha_v)\neq(\alpha_w)}$ is a product over all pairs of distinct vectors, (α_v) and (α_w), or for all pairs of treatments, $T(\alpha_v)$ and $T(\alpha_w)$. Reduction in (14) may be effected in view of (13). However, in order to obtain the maximum likelihood esti-

mators of parameters like α_{ri}, it is necessary to maximize (14) subject to the linear restraints, (12). The resulting equations have been examined and are difficult to solve. We have obtained an explicit solution for the 2×2 factorial but in other cases it appears that iterative solutions depending on successive approximations will be required. We have shown the general model to indicate that the theory may be carried through but we shall consider only the simple case in detail.

It is interesting to note that the model for factorials given here very closely resembles the ordinary analysis-of-variance model. We have already indicated, through the association of (7) and (8), that $\ln \pi_i$ may be regarded as a location parameter for T_i. In the same way, $\ln \pi(\alpha_r)$ is a location parameter for $T(\alpha_r)$ and from (13) we see that

$$\ln \pi(\alpha_r) = \sum_{i=1}^{m} \ln \alpha_{ri} . \tag{15}$$

Then, $\ln \alpha_{ri}$ is a parameter representing the effect of the r_ith level of factor i. The analogy is carried further, as noted above, for in both the analysis of variance and here there is one restraint introduced for the parameters of each factor. The restraints in both cases are introduced only to produce a determinate solution for estimators of the parameters and those restraints that we have chosen to facilitate our analysis are given in (12).

3. THE 2 × 2 FACTORIAL

For the 2×2 factorial it will be more convenient to replace the two factors, A_1 and A_2, in the general model, by factors, A and B. The levels of the factors, a_1 and a_2, both become two and we shall replace α_{12} and α_{22} by β_1 and β_2 while α_{11} and α_{21} become α_1 and α_2. Thus, we consider factors A and B, at two levels each, with parameters representing the effects of the levels of these factors given by α_1, α_2, β_1 and β_2 respectively. It is understood that $\alpha_i \geq 0$, $\beta_i \geq 0$, $i = 1, 2$, and the restraints similar to (12) are

$$\begin{aligned} \alpha_1 + \alpha_2 &= 1, \\ \beta_1 + \beta_2 &= 1. \end{aligned} \tag{16}$$

The four treatment combinations are designated by $T(\alpha_1 , \beta_1)$, $T(\alpha_1 , \beta_2)$, $T(\alpha_2 , \beta_1)$ and $T(\alpha_2 , \beta_2)$ and their parameters are respectively, $\pi(\alpha_1 , \beta_1) = \alpha_1\beta_1$, $\pi(\alpha_1 , \beta_2) = \alpha_1\beta_2$, $\pi(\alpha_2 , \beta_1) = \alpha_2\beta_1$, and $\pi(\alpha_2 , \beta_2) = \alpha_2\beta_2$. Substitution in (14) yields the likelihood function,

$$L(\alpha, \beta) = \alpha_1^{r_1}\alpha_2^{r_2}\beta_1^{t_1}\beta_2^{t_2}[(\alpha_1\beta_1 + \alpha_2\beta_2)(\alpha_1\beta_2 + \alpha_2\beta_1)]^{n} \cdot [(\alpha_1 + \alpha_2)(\beta_1 + \beta_2)]^{2n}, \tag{17}$$

with

$$
\begin{aligned}
c_1 &= 11n - \Sigma r(\alpha_1) \\
c_2 &= 11n - \Sigma r(\alpha_2) \\
d_1 &= 11n - \Sigma r(\beta_1) \\
d_2 &= 11n - \Sigma r(\beta_2).
\end{aligned} \tag{18}
$$

We use $\Sigma r(\alpha_1)$ to represent the total sum of ranks, for all comparisons and all n repetitions, of the two treatment combinations containing factor A at level one, and similar definitions follow for the other sums in (18). The value given for c_1 , for example, may be checked. α_1 appears in both the parameters, $\pi(\alpha_1 , \beta_1)$ and $\pi(\alpha_1 , \beta_2)$, and α_1 is a factor of the sums, $[\pi(\alpha_1 , \beta_1) + \pi(\alpha_1 , \beta_2)]$ and $[\pi(\alpha_1 , \beta_2) + \pi(\alpha_1 , \beta_1)]$, all of which correspond to terms in (14). The exponent of α_1 , c_1 , is equal to $\{2n(t - 1) - \Sigma r(\alpha_1 , \beta_1)\} + \{2n(t - 1) - \Sigma r(\alpha_1 , \beta_2)\} - n$. This reduces to the expression for c_1 given in (18) for now $t = 4$ and $\Sigma r(\alpha_1) = \Sigma r(\alpha_1 , \beta_1) + \Sigma r(\alpha_1 , \beta_2)$. $\Sigma r(\alpha_i , \beta_j)$ is the total sum of ranks assigned to $T(\alpha_i , \beta_j)$.

The following two relations also hold:

$$
\begin{aligned}
\Sigma r(\alpha_1) + \Sigma r(\alpha_2) &= \Sigma r(\beta_1) + \Sigma r(\beta_2) = 18n \\
c_1 + c_2 &= d_1 + d_2 = 4n.
\end{aligned} \tag{19}
$$

The second result follows from the first in view of (18), and it is also clear that $\Sigma r(\alpha_1) + \Sigma r(\alpha_2) = \Sigma r(\beta_1) + \Sigma r(\beta_2)$ because each sum represents the total of all ranks assigned in the experiment. That this total is $18n$ follows from the fact that there are $n\binom{4}{2} = 6n$ comparisons in the experiment and, since the 'preferred' or 'superior' treatment in each pair is given the rank one and the other the rank two, the sum of ranks for each comparison is three.

A number of hypotheses on the nature of the parameters α_1 , α_2 , β_1 , and β_2 will be imposed and used in the series of tests developed. For each test procedure, we must use approximate tests based on the distributions of the logarithms of the likelihood ratio statistics. We shall then consider first the estimation of the parameters by the method of maximum likelihood under these hypotheses and later show how they are used in tests.

Case (i): $H_1 : \alpha_1 = \alpha_2 , \beta_1 = \beta_2$.

This case is the most trivial one and the hypothesis specifies that all parameters have the value $\frac{1}{2}$. The hypothesis is equivalent to the assumption of no effects due to either the A or B factors. Substitution

in (17) is used and it follows that

$$-2 \ln L(\hat{\alpha}, \hat{\beta} \mid H_1) = 12n \ln 2. \tag{20}$$

We use $\hat{\alpha}$ and $\hat{\beta}$ to indicate that estimates obtained under the restrictions of the hypothesis have been substituted for the parameters in L. Case (ii): $H_2 : \alpha_1 = \alpha_2$, β_1 not assumed equal to β_2 .

$H_2' : \alpha_1$ not assumed equal to α_2 , $\beta_1 = \beta_2$.

Under H_2 , maximum likelihood estimates of β_1 and β_2 are easily obtained and are

$$\hat{\beta}_1 = \frac{d_1}{4n}, \qquad \hat{\beta}_2 = \frac{d_2}{4n}. \tag{21}$$

Then,

$$-2 \ln (\hat{\alpha}, \hat{\beta} \mid H_2) = 20n \ln 2 + 8n \ln n - 2d_1 \ln d_1 - 2d_2 \ln d_2 . \tag{22}$$

Similarly, under H_2' ,

$$\hat{\alpha}_1 = \frac{c_1}{4n}, \qquad \hat{\alpha}_2 = \frac{c_2}{4n} \tag{23}$$

and

$$-2 \ln (\hat{\alpha}, \hat{\beta} \mid H_2') = 20n \ln 2 + 8n \ln n - 2c_1 \ln c_1 - 2c_2 \ln c_2 . \tag{24}$$

Case (iii): $H_3 : \alpha_1$ not assumed equal to α_2 , β_1 not assumed equal to β_2 .

In this situation solution for maximum likelihood estimators of the parameters is difficult. Maximization of the likelihood function (17) using LaGrange multipliers, λ and μ, with the restraints (16) yields the set of equations,

$$\begin{aligned}
&\hat{\alpha}_1 + \hat{\alpha}_2 = 1, \\
&\hat{\beta}_1 + \hat{\beta}_2 = 1, \\
&\frac{c_1}{\hat{\alpha}_1} - \frac{2n}{\hat{\alpha}_1 + \hat{\alpha}_2} - \frac{n\hat{\beta}_1}{\hat{\alpha}_1\hat{\beta}_1 + \hat{\alpha}_2\hat{\beta}_2} - \frac{n\hat{\beta}_2}{\hat{\alpha}_1\hat{\beta}_2 + \hat{\alpha}_2\hat{\beta}_1} - \lambda = 0, \\
&\frac{c_2}{\hat{\alpha}_2} - \frac{2n}{\hat{\alpha}_1 + \hat{\alpha}_2} - \frac{n\hat{\beta}_2}{\hat{\alpha}_1\hat{\beta}_1 + \hat{\alpha}_2\hat{\beta}_2} - \frac{n\hat{\beta}_1}{\hat{\alpha}_1\hat{\beta}_2 + \hat{\alpha}_2\hat{\beta}_1} - \lambda = 0, \\
&\frac{d_1}{\hat{\beta}_1} - \frac{2n}{\hat{\beta}_1 + \hat{\beta}_2} - \frac{n\hat{\alpha}_1}{\hat{\alpha}_1\hat{\beta}_1 + \hat{\alpha}_2\hat{\beta}_2} - \frac{n\hat{\alpha}_2}{\hat{\alpha}_1\hat{\beta}_2 + \hat{\alpha}_2\hat{\beta}_1} - \mu = 0, \\
&\frac{d_2}{\hat{\beta}_2} - \frac{2n}{\hat{\beta}_1 + \hat{\beta}_2} - \frac{n\hat{\alpha}_2}{\hat{\alpha}_1\hat{\beta}_1 + \hat{\alpha}_2\hat{\beta}_2} - \frac{n\hat{\alpha}_1}{\hat{\alpha}_1\hat{\beta}_2 + \hat{\alpha}_2\hat{\beta}_1} - \mu = 0.
\end{aligned} \tag{25}$$

It is easy to show that $\lambda = 0$ and $\mu = 0$ and that solution of (25) may be obtained from the two independent equations,

$$(26)\qquad \begin{aligned} \frac{c_1}{\hat{\alpha}_1} - 2n - \frac{n\hat{\beta}_1}{1 - \hat{\alpha}_1 - \hat{\beta}_1 + 2\hat{\alpha}_1\hat{\beta}_1} - \frac{n(1 - \hat{\beta}_1)}{\hat{\alpha}_1 + \hat{\beta}_1 - 2\hat{\alpha}_1\hat{\beta}_1} &= 0, \\ \frac{d_1}{\hat{\beta}_1} - 2n - \frac{n\hat{\alpha}_1}{1 - \hat{\alpha}_1 - \hat{\beta}_1 + 2\hat{\alpha}_1\hat{\beta}_1} - \frac{n(1 - \hat{\alpha}_1)}{\hat{\alpha}_1 + \hat{\beta}_1 - 2\hat{\alpha}_1\hat{\beta}_1} &= 0. \end{aligned}$$

We transform by setting

$$(27)\qquad u = \hat{\alpha}_1 - \hat{\beta}_1 \quad \text{and} \quad v = \hat{\alpha}_1 + \hat{\beta}_1$$

and define

$$(28)\qquad x = \frac{c_1 + d_1 - n}{n} \quad \text{and} \quad w = \frac{c_1 - d_1}{n}.$$

The equations (26) yield a new set by multiplying the first by $\hat{\alpha}_1$ and the second by $\hat{\beta}_1$ and adding and subtracting the resultant forms. The transformation (27) is used and the variate u eliminated by noting that

$$(29)\qquad u^2 = v^2 - 2v + 3 + \frac{x - 3}{2v - x + 1}.$$

The equation for v is

$$\begin{aligned} 128v^5 - (160 + 160x)v^4 + (210 + 244x - 18w^2 + 66x^2)v^3 \\ + (61 - 271x - 117x^2 - 9x^3 - 9w^2 + 21w^2x)v^2 \\ + (4 - 54x + 120x^2 + 18x^3 - 8w^2x^2 + 6w^2x)v \\ + (w^2x^3 - w^2x^2 + 12x^2 - 18x^3 - 2x) = 0. \end{aligned} \qquad (30)$$

This equation may be solved for v using Newton's or Horner's method after isolation of the roots by Sturm's theorem. Values of v are bounded by 0 and 2. When v has been obtained, u is found using (29) and values of $\hat{\alpha}_1$ and $\hat{\beta}_1$ are available from (27). While this method of solution is somewhat clumsy, it is elementary and yields a simpler result than can be obtained by considering the set (26) directly.

If we assume that the solutions are $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\beta}_1$ and $\hat{\beta}_2$ and that they are known, then

$$\begin{aligned} -2 \ln L(\hat{\alpha}, \hat{\beta} \mid H_3) = 2n \ln (\hat{\alpha}_1\hat{\beta}_1 + \hat{\alpha}_2\hat{\beta}_2) + 2n \ln (\hat{\alpha}_1\hat{\beta}_2 + \hat{\alpha}_2\hat{\beta}_1) \\ - 2c_1 \ln \hat{\alpha}_1 - 2c_2 \ln \hat{\alpha}_2 - 2d_1 \ln \hat{\beta}_1 - 2d_2 \ln \hat{\beta}_2 . \end{aligned} \qquad (31)$$

Case (iv): $H_4 : \pi_{ij}$ not assumed equal to $\alpha_i\beta_j$.

This case is used for a test of interaction and H_4 is used to imply

that treatment parameters cannot be explained in terms of the α's and β's above. π_{ij} is used to denote the treatment rating for the treatment combination made up of the ith level of the A-factor and the jth level of the B-factor. In this case π_{11} , π_{12} , π_{21} and π_{22} correspond to π_1 , $\cdots$, π_4 and estimates are obtained as shown in the review of the general model with $t = 4$. Estimates may be used from tables in some cases and in any situation they may be calculated by solving the equations specified in (4, 5). The likelihood function is of the form (2) and, if p_{ij} is the estimator of π_{ij} ,

$$-2 \ln L(p_{ij} \mid H_4) = 2 \sum_{(ij)} [6n - \Sigma r(\alpha_i , \beta_j)] \ln p_{ij} - 2n \sum_{\substack{i \leq g, j \leq h \\ (i,j) \neq (g,h)}} \ln (p_{ij} + p_{gh}) \tag{32}$$

The first sum is over all treatments and the second over all treatment pairs.

4. TEST PROCEDURES FOR THE 2 × 2 FACTORIAL

The following tests are based on likelihood ratio statistics, $\lambda = L(\omega)/L(\Omega)$, where $L(\omega)$ is the maximum of the likelihood function under the restrictions of a null hypothesis, H_0 , and $L(\Omega)$ is the corresponding function under H_a . Under quite general conditions $-2 \ln \lambda$ has a χ^2-distribution for large samples. We have shown the forms, $-2 \ln L$, for a number of hypotheses and these may be quickly used to obtain the appropriate statistics and their approximate distributions.

I. An over-all test for treatment effects assuming no interaction.

$H_0 = H_1 : \alpha_1 = \alpha_2 , \beta_1 = \beta_2 .$

$H_a = H_3 : \pi_{ij} = \alpha_i\beta_j$; $\alpha_1 \neq \alpha_2$ or $\beta_1 \neq \beta_2$ or both.

The test statistic is

$$-2 \ln \lambda_{13} = 12n \ln 2 + 2c_1 \ln \hat{\alpha}_1 + 2c_2 \ln \hat{\alpha}_2 + 2d_1 \ln \hat{\beta}_1 + 2d_2 \ln \hat{\beta}_2 - 2n \ln (\hat{\alpha}_1\hat{\beta}_1 + \hat{\alpha}_2\hat{\beta}_2) - 2n \ln (\hat{\alpha}_1\hat{\beta}_2 + \hat{\alpha}_2\hat{\beta}_1), \tag{33}$$

and for large samples it has the distribution of χ^2 with 2 degrees of freedom. The estimators in (33) are those specified under Case (iii).

II. An over-all test for treatment effects admitting possible interaction.

$H_0 : \pi_{ij} = \frac{1}{4}$ for all i and j. (equivalent to H_1).

$H_a = H_4 : \pi_{ij}$ possibly unequal to $\alpha_i\beta_j$ for some i and j.

This is the general method of paired comparisons reviewed in the introduction. The test statistic is

$$-2 \ln \lambda_{14} = 12n \ln 2 + 2 \sum_{(ij)} [6n - \Sigma r(\alpha_i , \beta_j)] \ln p_{ij} - 2n \sum_{\substack{i \leq g , j \leq h \\ (i,j) \neq (g,h)}} \ln (p_{ij} + p_{gh}) \tag{34}$$

and its distribution approximates that of χ^2 with 3 degrees of freedom when n is large. The exact distribution of B_1 , a monotone function of $-2 \ln \lambda_{14}$, has been tabled for values of n up to 8.

III. A test of $A \times B$ interaction.

$$H_0 = H_3 : \pi_{ij} = \alpha_i \beta_j \ ; \alpha_1 \neq \alpha_2 \text{ or } \beta_1 \neq \beta_2 \text{ or both.}$$

$$H_a = H_4 : \pi_{ij} \neq \alpha_i \beta_j \text{ for some } i \text{ and } j.$$

This test is one of whether the data can be explained in terms of parameters α and β or whether the slightly more flexible parameters π_{ij} are required. It amounts to a test of $A \times B$ interaction and the statistic required is

$$-2 \ln \lambda_{34} = -2 (\ln \lambda_{14} - \ln \lambda_{13}). \tag{35}$$

This statistic approximates χ^2 with 1 degree of freedom.

IV. Tests on one factor assuming the non-existence of the other.

$$H_0 = H_1 : \quad \alpha_1 = \alpha_2 , \quad \beta_1 = \beta_2 ,$$

$$H_a = H_2 : \quad \alpha_1 = \alpha_2 , \quad \beta_1 \neq \beta_2 .$$

The test statistic for the B-factor is

$$-2 \ln \lambda_{12} = 2d_1 \ln d_1 + 2d_2 \ln d_2 - 8n \ln n - 8n \ln 2. \tag{36}$$

Similarly, for the A-factor,

$$H_0 = H_1 : \quad \alpha_1 = \alpha_2 , \quad \beta_1 = \beta_2 ,$$

$$H_a = H_2' : \quad \alpha_1 \neq \alpha_2 , \quad \beta_1 = \beta_2 .$$

The test statistic is

$$-2 \ln \lambda_{12'} = 2c_1 \ln c_1 + 2c_2 \ln c_2 - 8n \ln n - 8n \ln 2. \tag{37}$$

Both $-2 \ln \lambda_{12}$ and $-2 \ln \lambda_{12'}$ have the distribution of χ^2 with one degree of freedom for large samples.

V. Tests of one factor without regard to the other.

$$H_0 = H_2 : \alpha_1 = \alpha_2 , \beta_1 \text{ not assumed equal to } \beta_2 .$$

$$H_a = H_3 : \pi_{ij} = \alpha_i \beta_j \ ; \alpha_1 \neq \alpha_2 \text{ or } \beta_1 \neq \beta_2 \text{ or both.}$$

or

$$H_0 = H_2' : \beta_1 = \beta_2 \text{ , } \alpha_1 \text{ not assumed equal to } \alpha_2 \text{ .}$$

$$H_a = H_3 : \pi_{ij} = \alpha_i\beta_j \text{ ; } \alpha_1 \neq \alpha_2 \text{ or } \beta_1 \neq \beta_2 \text{ or both.}$$

The first statistic, for the B-factor, is

$$-2 \ln \lambda_{23} = -2(\ln \lambda_{13} - \ln \lambda_{12}) \tag{38}$$

and the second, for the A-factor, is

$$-2 \ln \lambda_{2\cdot 3} = -2(\ln \lambda_{13} - \ln \lambda_{12\cdot}). \tag{39}$$

Again, both statistics have the distribution of χ^2 with one degree of freedom for large samples.

A summary of these test procedures is given in Table I. In the table we show how the tests are related and it is interesting to note that the over-all test for treatment effects assuming no interaction may be partitioned in two ways in a fashion somewhat similar to blocks and treatments in an incomplete block design. In the latter situation the analysis may show either blocks (unadjusted) and treatments (adjusted) or blocks (adjusted) and treatments (unadjusted).

In the introductory section we showed that the basic paired comparison experiment may be considered in g groups of repetitions and that a test for interaction of treatments with groups could be obtained. For the 2 × 2 factorial, it does not seem possible to get a measure of a treatment effect by group interaction directly as a likelihood ratio test. Tests within groups can be obtained simply by adding the appropriate statistics computed for each group over all of the groups. For example, $A \times B$ interaction within groups is measured by the sum,

$$-2 \sum_{u=1}^{g} \ln \lambda_{34.u}$$

where $-2 \ln \lambda_{34.u}$ is computed using (35) and consideration is restricted to the uth of g groups of repetitions. This composite statistic has the distribution of χ^2 with g degrees of freedom for large samples. Further study will be given to the problem of measuring and testing interactions like that of $A \times B \times$ groups but this interaction is not isolated in the analysis of variance for balanced incomplete block designs.

5. AN ILLUSTRATIVE EXAMPLE

Suppose that the sums of ranks for the four treatment combinations, A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 , are

$$\Sigma r(\alpha_1 \text{ , } \beta_1) = 38, \qquad \Sigma r(\alpha_2 \text{ , } \beta_1) = 48,$$

$$\Sigma r(\alpha_1 \text{ , } \beta_2) = 44, \qquad \Sigma r(\alpha_2 \text{ , } \beta_2) = 50,$$

TABLE I

TEST PROCEDURES FOR THE 2 × 2 FACTORIAL

Test	Hypotheses	Statistic	Degrees of freedom	Computing references
Treatment effects assuming no interaction	$H_0 = H_1 : \alpha_1 = \alpha_2 , \beta_1 = \beta_2 .$ $H_a = H_3 : \pi_{ij} = \alpha_i\beta_j ; \alpha_1 \neq \alpha_2$ or $\beta_1 \neq \beta_2$ or both.	$-2 \ln \lambda_{13}$	2	(33).
A-factor	$H_0 = H_2 : \alpha_1 = \alpha_2 , \beta_1$ not assumed equal to $\beta_2 .$ $H_a = H_3 : \pi_{ij} = \alpha_i\beta_j ; \alpha_1 \neq \alpha_2$ or $\beta_1 \neq \beta_2$ or both.	$-2 \ln \lambda_{23} = -2(\ln \lambda_{13} - \ln \lambda_{12}).$	1	(39), (33), (36).
B-factor assuming no A-factor effect	$H_0 = H_1 : \alpha_1 = \alpha_2 , \beta_1 = \beta_2 .$ $H_a = H_2 : \alpha_1 = \alpha_2 , \beta_1 \neq \beta_2 .$	$-2 \ln \lambda_{12}$	1	(36).
B-factor	$H_0 = H_{2'} : \beta_1 = \beta_2 , \alpha_1$ not assumed equal to $\alpha_2 .$ $H_a = H_3 : \pi_{ij} = \alpha_i\beta_j ; \alpha_1 \neq \alpha_2$ or $\beta_1 \neq \beta_2$ or both.	$-2 \ln \lambda_{2'3} = -2(\ln \lambda_{13} - \ln \lambda_{12'})$	1	(38), (33), (37).
A-factor assuming no B-factor effect	$H_0 = H_1 : \alpha_1 = \alpha_2 , \beta_1 = \beta_2 .$ $H_a = H_{2'} : \alpha_1 \neq \alpha_2 , \beta_1 = \beta_2 .$	$-2 \ln \lambda_{12'}$	1	(37).
$A \times B$ interaction	$H_0 = H_3 : \pi_{ij} = \alpha_i\beta_j ; \alpha_1 \neq \alpha_2$ or $\beta_1 \neq \beta_2$ or both. $H_a = H_4 : \pi_{ij} \neq \alpha_i\beta_j$ for some i and j.	$-2 \ln \lambda_{34} = -2(\ln \lambda_{14} - \ln \lambda_{13})$	1	(35), (34), (33).
Treatment effects admitting interaction	$H_0 = H_1 : \pi_{ij} = \frac{1}{4}.$ $H_a = H_4 : \pi_{ij}$ possibly unequal to $\alpha_i\beta_j$ for some i and j.	$-2 \ln \lambda_{14}$	3	(34).

with $n = 10$. We require

$$c_1 = 28, \quad d_1 = 24, \tag{40}$$
$$c_2 = 12, \quad d_2 = 16,$$

using (18). Solution of equations (4, 5) for p_{11} , p_{12} , p_{21} , p_{22} yields

$$p_{11} = .4729, \quad p_{21} = .1580, \tag{41}$$
$$p_{12} = .2423, \quad p_{22} = .1268.$$

Substitution in (28) gives us

$$x = 4.2 \quad \text{and} \quad w = 0.4. \tag{42}$$

Equation (30) becomes

$$128v^5 - 832v^4 + 2396.16v^3 - 3795.2v^2 + 3209.0368v - 1121.27232 = 0.$$

The root, obtained by Newton's method, is

$$v = 1.31275$$

from whence, by (29),

$$u = .095077.$$

Having found u and v, we may use (27) to obtain

$$\hat{\alpha}_1 = .7039, \quad \hat{\beta}_1 = .6088, \tag{43}$$
$$\hat{\alpha}_2 = .2961, \quad \hat{\beta}_2 = .3912.$$

We now have those quantities required to evaluate the statistics in Table I.

We first compute the two statistics, $-2 \ln \lambda_{13}$ and $-2 \ln \lambda_{14}$, for treatment effects. We show the computation for brevity in terms of natural logarithms although it may facilitate computing to use common logarithms and a conversion factor. To compute $-2 \ln \lambda_{13}$, we use the estimates in (43) and substitute in (33). Then

$$\begin{aligned} -2 \ln \lambda_{13} = {} & 120 \ln 2 + 56 \ln .7039 + 24 \ln .2961 \\ & + 48 \ln .6088 + 32 \ln .3912 \\ & - 20 \ln [(.7039)(.6088) + (.2961)(.3912)] \\ & - 20 \ln [(.7039)(.3912) + (.2961)(.6088)] \\ = {} & 8.3356. \end{aligned}$$

TABLE II

THE RANK ANALYSIS FOR A 2 × 2 FACTORIAL WITH $n = 10$

Test	Degrees of freedom	χ^2	Significance level
Treatment effects assuming no interaction	2	8.3356*	.016
A-factor	1	6.7249**	.009
B-factor assuming no *A* effect	1	1.6107	.205
B-factor	1	1.7528	.190
A-factor assuming no *B* effect	1	6.5828**	.010
$A \times B$ interaction	1	0.4492	.504
Treatment effects admitting interaction	3	8.7848	.034

From (34) with the estimates (41),

$$-2 \ln \lambda_{14} = 120 \ln 2 + 44 \ln .4729 + 32 \ln .2423 + 24 \ln .1580 + 20 \ln .1268 - 20 \ln .7152 - 20 \ln .6309 - 20 \ln .5997 - 20 \ln .4003 - 20 \ln .3691 - 20 \ln .2848 = 8.7848.$$

The $A \times B$ interaction effect is obtained by subtraction in view of (35) and

$$-2 \ln \lambda_{34} = 8.7848 - 8.3356 = 0.4492.$$

It is now only necessary to compute the two test statistics for the effects of each factor assuming that the other does not exist as given in IV. For the *B*-factor,

$$-2 \ln \lambda_{12} = 48 \ln 24 + 32 \ln 16 - 80 \ln 10 - 80 \ln 2 = 1.6107,$$

and, for the *A*-factor,

$$-2 \ln \lambda_{12'} = 56 \ln 28 + 24 \ln 12 - 80 \ln 10 - 80 \ln 2 = 6.5828.$$

For a test of the *A*-factor with no assumptions on the *B*-factor, we have, by subtraction,

$$-2 \ln \lambda_{23} = 8.3356 - 1.6107 = 6.7249$$

and, for the B-factor without assumptions on the A-factor,

$$-2 \ln \lambda_{2'3} = 8.3356 - 6.5828 = 1.7528.$$

The analysis is summarized in Table II. We see that the interaction effect is non-significant, the B-factor effect is non-significant, but the A-factor effect is significant at the .01 level.

6. DISCUSSION

When a new method of analysis is presented, it is desirable to compare it with possible alternative techniques. Some preliminary work has been done [6] in investigating the power of the general method of paired comparisons. In that case, it was found that the asymptotic efficiency of the method of paired comparisons relative to analysis of variance, and under those conditions appropriate for the valid application of analysis of variance, is $t/\pi(t-1)$. When $t = 2$, this reduces to $2/\pi$, the relative efficiency of the sign test. For large numbers of treatments, this efficiency approaches $1/\pi$. Then, on the basis of the number of samples required for each treatment, the ratio of sample numbers required for paired comparisons as contrasted with analysis of variance approaches some number on the approximate range, $\frac{3}{2}$ to 3. However, this is an approximate result, and Dixon [8] has recently shown that the sign test may compare much more favorably with analysis of variance for small sample sizes than indicated by the result on relative efficiency. Further, the method of paired comparisons may be used in many situations where it is difficult or invalid to use analysis of variance, and the ease of making paired comparisons may alone often justify the use of the method.

The powers of the test procedures set forth for the 2 × 2 factorial arrangement in paired comparisons have not been investigated. It does seem reasonable however to expect that they conform quite closely to the result here stated for the general method and thus to the sign test.

The analysis of the 2 × 2 factorial in paired comparisons in some ways resembles factorial chi-square analyses [7]. There are some points of difference, and it is not believed that the methods here presented represent an extension of factorial chi-square to incomplete blocks. Brandt illustrates the factorial chi-square method with a 2 × 3 × 3 factorial experiment wherein he records only the frequencies of success or failure of given treatment combinations. We could represent the event that a treatment obtains rank one in a paired comparison as a success and then record, for each treatment combination, the frequency of success or of rank one. Our data could then be put in a

table similar to those used by Brandt but intercorrelations between these frequencies would then enter. Under the model for paired comparisons, the expected frequency of success for each treatment would depend on the values of all of the treatment parameters in the experiment and this would be a much more complicated expression than expected in factorial chi-square.

We have presented a rank analysis for 2×2 factorial treatments in the method of paired comparisons. The theory could be developed for general factorial sets of treatments but the solutions of estimating equations present formidable difficulties. The 2×2 factorial alone offers interesting possibilities in the designing of taste testing experiments, where paired comparisons have proved particularly appropriate, and other applications may be expected.

It is believed that this is the first time that factorial treatments have been considered with ranked data and it is noteworthy to find that the usual comparisons of the analysis of variance may be made.

A more detailed discussion of the material presented here is given in the thesis [1].

REFERENCES

1. Abelson, R. M. A Ranking Experiment with Paired Comparisons and a Factorial Design. M.S. Thesis, Virginia Polytechnic Institute, Blacksburg, 1953.
2. Bradley, R. A. and Terry, M. E. The Rank Analysis of Incomplete Block Designs. I The Method of Paired Comparisons. *Biometrika 39*, 324–345, 1952.
3. Bradley, R. A. Incomplete Block Rank Analysis: On the Appropriateness of the Model for a Method of Paired Comparisons. *Biometrics 10*, 375–391, 1954.
4. Bradley, R. A. Some Statistical Methods in Taste Testing and Quality Evaluation. *Biometrics 9*, 22–38, 1953.
5. Bradley, R. A. The Rank Analysis of Incomplete Block Designs. II Additional Tables for the Method of Paired Comparisons. *Biometrika 41*, 319–354, 1954.
6. Bradley, R. A. Some Large-Sample Results on Estimation and Power for a Method of Paired Comparisons. (Preliminary Report). Abstract, *Annals of Math. Stat. 24*, 688, 1953.
7. Brandt, A. E. Use of Factorial Chi-Square in Test Shooting. Unpublished manuscript.
8. Dixon, W. J. Power Functions of the Sign Test and Power Efficiency for Normal Alternatives. *Annals of Math. Stat. 24*, 467–473, 1953.
9. Hopkins, J. W. Incomplete Block Rank Analysis: Some Taste Test Results. *Biometrics 10*, 391–402, 1954.
10. Mosteller, F. Remarks on the Method of Paired Comparisons. I. The Least Squares Solution Assuming Equal Standard Deviations and Equal Correlations. *Psychometrika 16*, 3–9, 1951.
11. Thurstone, L. L. Psychophysical Analysis. *Amer. J. Psychology 38*, 368–389, 1927.

SOME CONSIDERATIONS IN THE DESIGN OF SUCCESSIVE EXPERIMENTS IN FRUIT PLANTATIONS*

T. N. Hoblyn, S. C. Pearce and G. H. Freeman

East Malling Research Station, Kent

PART I. GENERAL

The designer of an experiment on perennial plants, especially on trees, must take into account several considerations which do not ordinarily concern those whose crops grow and mature in a single season, namely (1) those that affect any initial trial, (2) those that affect any subsequent trial and (3) the desiderata in an initial trial to ensure that it shall be possible to arrange subsequent trials on the same material.

The Initial Trial

First, the experiment is usually unique. Only occasionally can a trial be repeated either in situation or in more than one series of seasons; even then exact repetition is seldom possible, so that each trial must be designed to yield results of adequate precision on its own. For this reason alone, if the design is at all complex, there must be a very clear appreciation of the issues under investigation and the plots must be replicated to a greater extent than may be necessary when the trial is one of a series. Sometimes, as in a spraying trial, repetition in time is possible, but if the same trees are used in successive years without change of treatments the cumulative effects of previous seasons, not only on growth or cropping but also on the incidence of disease or pests, must always be taken into account.

Secondly, because of their size, there are fewer plants per plot than in trials of cereals or vegetables. Thus both the history of the tree prior to transplantation to its permanent position in the field and local differences in its environment after that date may be of great importance in determining the experimental error. With some crops genetical differences in the plant itself or in its several parts (e.g., the rootstock

*Part I contains the substance of a paper by T. N. Hoblyn and S. C. Pearce read at the Third International Biometric Conference. The Red Spider trial in Part II was the subject of an exhibit in the names of S. C. Pearce and G. H. Freeman on the same occasion.

and the scion of a fruit tree) are important, though the use of clonal, i.e., vegetatively propagated, material reduces this source of variation considerably. Even where such material is available, however, small differences in treatment in the nursery, during transplanting and in the early training during establishment may all contribute towards tree to tree variation. Thus the eventual performance of an individual tree or plot is usually a complex of the effects of its history and its environment, each of which may contribute substantially to its behaviour

Where an experiment is one that starts with the propagation of the plants, e.g., variety or rootstock trials, or as soon as they are transplanted, e.g., trials on methods of tree formation, there is no way of reducing experimental error other than by extreme care in the cultural operations. If, however, the start of a trial can be delayed until the trees are established in the field (e.g., trials of cultural, manurial or disease control operations on the maturing tree) something can be done to reduce the error by the device of calibration (Pearce and Taylor, 1950) in which the analysis of covariance is used to correct the performance of individual trees after the experiment has started by means of their recorded pre-history. In such instances it may be possible to reduce the error variance by as much as one-third.

Subsequent Trials

Experiments on trees usually last for a long period, but not by any means all of them need continue throughout the whole life of a plantation. Few experiment stations can afford to remove a plantation of trees on which the initial experiment has been completed while the trees are yet in their prime; and it is, therefore, as well for the statistician to make provision in his design for new unspecified uses for the plantation after its original purpose has been fulfilled Apart from this, it is impossible to provide a constant supply of unused trees to meet demands for experiments that can only be conducted in a mature orchard, so that experimental orchards designed for a succession of experiments are an undoubted asset at any horticultural research institute. Indeed, established trees are of especial value because they may respond to treatments quite differently from young ones, and it is the behaviour of the older trees that is often of greater interest.

Before proceeding further it will be as well to consider the assumptions implicit in this sort of experimentation:

(1) The first experiment should have left the plants normal as to both present condition and growth, so that they remain acceptable experimental material. Sometimes one or more treatments will have

led to abnormal trees, which will have to be rejected from a further trial; if all treatments have done this, then all trees must be rejected, at least until after remedial treatment, despite their similarity one to another.

(2) There should be no interaction between original and subsequent treatments. In practice it is rarely possible to be certain that this is so, but an interaction is often unlikely and the consequence of there being one is loss of precision not lack of validity. In order to ensure that this assumption holds various devices are used. (a) A change may be made to treatments quite different from those previously applied, thus, a former rootstock trial may be used for investigating methods of pest control. (b) Trees are rejected if they are very different from the rest on account of the previous treatments. (c) In spraying trials it is usually sufficient to be sure that none of the original spray materials still adhere to the plants.

All these points need to be considered from the widest biological standpoint; thus, at first sight it might seem that rootstocks are indeed unlikely to interact with methods of pest control on the scion. However, they might do so with sprays around blossoming time, because the date of blossom can be slightly affected by the rootstock.

(3) With repeated changes, when the $(n + 2)$th set of treatments is applied there should no longer be any residual effects from more than one of the former $(n + 1)$ sets, the other n having been eliminated by the course of time. This brings up the whole question of persistence of treatments. Some cannot be eliminated; others can be relied upon with some certainty to disappear. In practice, repeated changing takes place principally with spraying trials, and here persistence can be due to four causes: (a) Physical damage to the trees by some of the treatments, e.g., by scorching the leaves. Such effects usually disappear fairly quickly but not if the damage has been severe. (b) Chemical residues on the leaves, fruits and bark, which are usually removed fairly rapidly by the weather; in any event, in the autumn the fruit will be picked and the leaves will fall. Systemic poisons, however, sometimes persist for a surprisingly long time. (c) Uneven distribution of the pest or pathogen on account of previous treatments. However, almost all species have some means of motion and at some time of the year disperse anyway. A point to be noted here is that good control of an insect pest may prevent its becoming overcrowded in the summer and this may interfere with its normal migration habits. This, of course, does not prevent an influx at migration time if the host species is otherwise common locally. (d) Persistence of a differential population of predators even after the species under study has dispersed. However,

predatory species are often mobile and tend to follow their prey, though not always immediately.

It should be emphasized that no residual effect should be assumed to have disappeared unless there are biological reasons for expecting it to have done so, and also an inspection of current data gives no cause for suspecting the unusual. The subject certainly requires careful thought, but after a little experience with a species need not cause much anxiety.

The problem as a whole has been discussed previously (Pearce and Taylor, 1948). In this paper it is intended to indicate some of the available arrangements when it can be assumed that the new treatments will not interact with those previously applied. Such designs are more easily applied if the need for them has been foreseen at the start.

Possible Designs for Subsequent Trials

One preliminary point should be noted. As has already been emphasized, positional variation is not of the same relative importance with perennials as it is with annuals. Consequently there is not the same objection to a treatment being represented more than once in a block, and devices such as balancing admit of a useful extension, namely, the addition to a block of complete replicates of the treatments over and above those needed for balance. (Thus, since this is a possible lay-out,

Block I,	ABC.	Block III,	BCD.
Block II,	ABD.	Block IV,	ACD.,

so also is

Block I,	AABBCCD.	Block III,	ABBCCDD.
Block II,	AABBCDD.	Block IV,	AABCCDD.)

Given an existing design four ways will be considered of allotting the new treatments to the blocks:—

O or orthogonally and with equal replication,
T or balanced totally in the sense of Yates (1936a),
P or balanced partially in the sense of Bose and Nair (1939),
and S or in supplemented balance.

In this last device a group of treatments are equally replicated and arranged so that any pair of them occur together in the same number of blocks, as in the following example, where each pair of A, B, C and D occur together four times. In addition there is a supplementing treatment, usually the Control, which occurs equally frequently with

each of the main treatments. Thus, each of A, B, C and D occurs three times (not four times) with the Control (O).

Block I,	ABCO.	Block IV,	BCDO.
Block II,	ABDO.	Block V,	ABCD.
Block III,	ABCD.	Block VI,	ACDO.

If one treatment of a pair occurs twice in one block, as may happen, and the other once, that counts as two occurrences together, not one. It will be noted that this device differs from the others considered here in that all treatments are not equally replicated.

Similar ways apply of disposing the new treatments with respect to plots of the original ones.

In the notation to be adopted the first letter will indicate the manner in which the original treatments are disposed with regard to blocks, and then, after a colon, will be shown in order how the new treatments are disposed with regard to the blocks and with regard to the original treatments. Thus a design of Type O:PT is one in which the original treatments were orthogonal to blocks and to which new treatments have been applied that are partially balanced with respect to blocks and totally balanced with respect to the plots of the original treatments. It is not proposed to discuss all possible designs here but principally those that can be used when the original design is in randomized blocks.

Simplest are those of Type O:OO, which are based on Latin squares, multiple if necessary, with blocks and initial treatments instead of rows and columns, but those of most general availability are those of Types O:OT and O:TO, which are corresponding analogues of Youden squares. The only design of Type O:TT of reasonable size is that given by Pearce and Taylor (1948). It is, however, sometimes very useful.

Partial balance does not in practice find much favour, because some treatment differences are compared more accurately than others, and this is undesirable except when, as in supplemented balance, there is only one odd treatment, which can be the control. This is a disadvantage of designs of Types O:OP, O:PT, etc.; some designs of Type O:PP, however, in which the pairs of new treatments occurring together least frequently in the same block coincide most frequently on the same original treatments and vice versa, largely avoid the criticism. For the designs of Type O:PP which are of greatest practical interest, the method of analysis is less complicated than might be supposed; for other O:PP and for O:TP designs, however, the analysis is rather more complicated, though never unworkably so.

Supplemented balance is a very valuable device, especially in a certain group of designs of Type O:SS that are the analogues of Latin

squares with (a) a row and a column missing (Yates, 1936b), or (b) a row and a column added, the supplementing treatment being applied to the vacant plot where the added row and column cross, or (c) a row added and a column missing (Pearce, 1952). These have the advantage that the replication of the new treatments is about the same as that of the initial ones, a very important point in practice, for it is rarely expedient to change one set of, say, three treatments with eight replicates for another of eight treatments with only three replicates. Either the one has been too much or the other will prove too little.

Designs to Facilitate Subsequent Changes

From what has been said, it will appear that the most generally useful designs are those of Types O:OO, O:OT, O:TO and O:SS. What does this imply in the design of the initial trial if it is expected that the trees will last longer than the original experiment? First of all, it means that the most convenient design is one in which the blocks and original treatments are equally numerous. Then if the next set of treatments number as many as the first set, a simple design of Type O:OO will be possible, and if they number one more or one fewer, Type O:SS will be possible. For the reason given, namely, the need for a similar degree of replication in the two experiments, a greater disparity in the numbers of treatments is unlikely to be required.

If it is not possible to arrange for such an initial design, the next best is to have the numbers of blocks and original treatments differing by only one. Then, if the new treatments are equal in number to the blocks, Type O:OT is possible; if they are equal in number to the first set of treatments, Type O:TO is available. Either way, a satisfactory design results.

Sometimes the possibility of further changes must be considered and, in the second case just mentioned, it is especially easy when treatments are applied for only a short time and do not give lasting residual effects, as happens, for example, with some spraying trials against pests and diseases. A design of Type O:OT, such as those just described, when the effects of the original treatments have disappeared, degenerates to one of Type O, which can again become either Type O:OT or O:TO. In similar circumstances a design of Type O:TO degenerates to one of Type T, which with the addition of further treatments can become Type T:OO or T:TT. In fact, by always using sets containing treatments to the number of either the blocks or the first set of treatments applied, the changing of treatments can continue indefinitely, a new set being applied as soon as the residual effect of the last set but one has disappeared.

PART II. EXAMPLES OF CHANGING TREATMENTS

When efficiency factors are given in the designs below, they refer to the final set of treatments in each case. The efficiency factor of one set of treatments is frequently altered by the subsequent application of another set, but, in general, the original set will no longer be of sufficient interest to be specially considered.

Single Changes of Treatments

Orthogonal designs. As has been mentioned in Part I the simplest method of adding a fresh set of treatments is that available when the numbers of blocks, of original treatments and of new treatments are all equal, for then a Latin square type of design is possible (Type O:OO). This is illustrated in Fig. 1 showing a trial for the control of

D F C A B E	B F E A D C	A F D E B C	A B F E C D	E B D A F C	C A D E B F
e d a b c f	a f e c d b	a b f d e c	e d c a f b	c b a f e d	e d c b f a

FIG. 1. EXAMPLE OF CHANGING TREATMENTS BY CONVERTING A DESIGN IN RANDOMIZED BLOCKS INTO THE ANALOGUE OF A LATIN SQUARE.

Key A-F 1951 Treatments. a-f 1952 Treatments.

Apple Mildew, *Podosphaera leucotricha* (Ell. and Everh.) Salm., in which a row of M.III rootstocks, 240 feet long, was divided into 36 plots of equal length. In 1951 six spraying treatments, represented by capital letters, were applied in a randomized block design; and in 1952 six different spraying treatments, represented by small letters, were applied instead. An analysis of variance on the second year's figures is possible, in which blocks and original treatments are dealt with as if they were the rows and columns of a Latin square.

Designs using total balance. A design such as the above, though admirable, is not generally available, and, more commonly, it is necessary to use designs of Types O:OT and O:TO. Thus, Fig. 2 illustrates a spraying trial against Cherry Scab, *Fusicladium Cerasi* (Rabenh.) Sacc., in which recourse was had to total balance. The material was an old rootstock trial or collection (the distinction between the two was hardly made in the days when it was planted) in two parts with a considerable overlap between the rootstocks represented in them. After rejection of unsatisfactory trees there were twelve rootstocks common to the two parts (or blocks) and the four spraying treatments were applied so that each occurred three times in each block and so that each combination of two treatments (i.e., ab, ac, ad, bc, bd and cd) occurred on two rootstocks. The design was thus of Type O:OT. The efficiency factor as compared with an orthogonal design

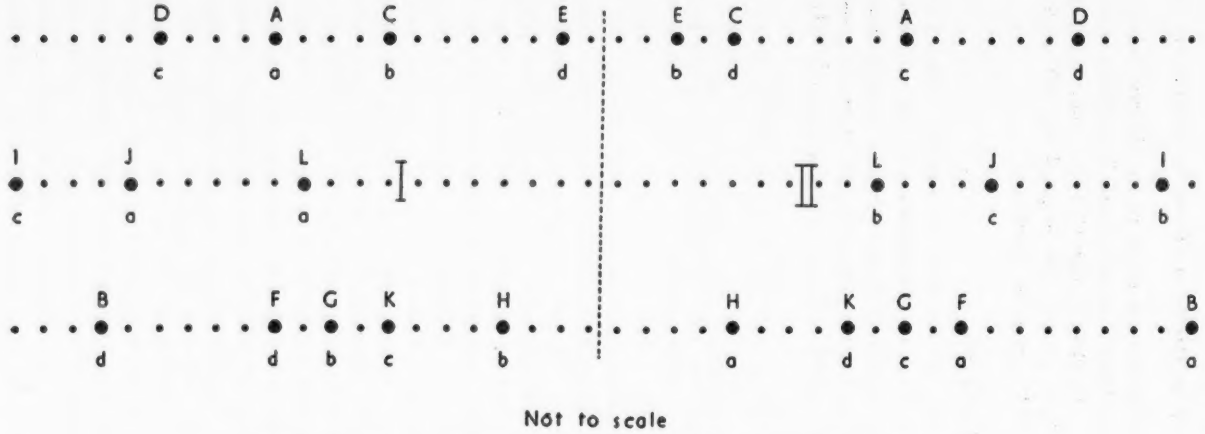


FIG. 2. EXAMPLE OF CHANGING TREATMENTS USING A DESIGN OF TYPE O:OT.
Key—A–L Original rootstocks. *a–d* Subsequent spraying treatments.

was 2/3, which is rather low. The systematic arrangement of the original rootstocks would not affect the comparison of the subsequent spraying treatments.

A similar design was used in an investigation in which five treatments, namely, the injections of four viruses and a control, were tested on another former cherry rootstock trial. After examination of the plantation it was decided to use forty trees of Turkey Heart on ten rootstocks in four randomized blocks, and to arrange that each new treatment occurred twice in each block. Also, each injection treatment was omitted from two rootstocks, so the design was again of Type O:OT, the efficiency factor this time being 15/16. This is a more usual figure, the loss of efficiency only rarely being serious.

In both these cherry trials the designer has adapted the original trial, at least in some measure, to fit the new one, instead of the other way round. In practice this must happen sometimes, especially when changing the treatments in a trial not originally designed with that purpose in view; but modern experiments on trees should be designed as far as possible to avoid this necessity.

Possibilities using partial balance. Up to the present at East Malling it has been possible to avoid partial balance, but many designs exist that could be used despite the objections mentioned in Part I. It may be of interest, therefore, to indicate some possible designs.

Designs of Types O:OP and O:PO are very easily constructed. It is not perhaps so obvious that designs of Type O:PP are fairly common, such as, for example, the one illustrated in Fig. 3. Here, it will

Block I	Aa	Bb	Cj	Di	Ee	Fh
Block II	Ac	Bg	Ca	Dd	Eh	Fj
Block III	Ae	Ba	Cf	Dg	Ei	Fd
Block IV	Ad	Bh	Ci	Db	Ec	Ff
Block V	Ab	Bf	Ce	Dj	Eg	Fe

FIG. 3. SCHEME FOR CHANGING TREATMENTS USING A DESIGN OF TYPE O:PP.
Key A-F First set of treatments. a-j Second set of treatments.

be seen, the comparisons of new treatments are of two types: (i) those like a and b, which occur together in only one block, namely I, but on two original treatments, namely A and B, and (ii) those like a and d, which occur together in two blocks, namely II and III, but on only one original treatment, namely A. The efficiency factors for these two sorts of comparison are very similar, namely 60/73 and 60/74 respectively.

Designs of Types O:PT and O:TP are also reasonably common, one being illustrated in Fig. 4. Several have been suggested by Shrik-

Block I	Aa	Bd	Cg	Da	Eb	Fc	Gd	He	If	Jg	Kh	Li
Block II	Ab	Be	Ch	Dd	Ee	Ff	Gi	Hg	Ih	Jb	Kc	La
Block III	Ac	Bf	Ci	Dg	Eh	Fi	Gb	Hc	Ia	Jf	Kd	Le

FIG. 4. SCHEME FOR CHANGING TREATMENTS USING A DESIGN OF TYPE O:PT.
Key A-L First set of treatments. a-i Second set of treatments.

hande (1951). In Fig. 4 Block I contains a complete set of new treatments together with a, d and g. In Block II the additional treatments are b, e and h and in Block III c, f and i, the additional treatments being partially balanced in respect of blocks. Also, the new treatments are totally balanced in respect of the original ones, so the design is of Type O:PT. Comparisons like that of a with d, which occur as additional treatments in the same block, have an efficiency factor of 3/4; while for those like a and b it is 99/136 (= 0.73).

In general, if these designs have few plots the efficiency is rather unsatisfactory. Thus, in the one just discussed, the new treatments are compared with nearly the same accuracy as were the original ones, although each is replicated four times instead of three. Nevertheless, although such designs have certain disadvantages, the writers believe them to have a useful rôle in research, providing as they do resources for changing treatments where other and simpler methods are not available.

Repeated Changes of Treatments

Orthogonal designs. In the trial shown in Fig. 1, if the residual effect of the first set of treatments should disappear, it would be possible to add a third set, the blocks and second set of treatments representing the rows and columns of a further Latin square.

In practice care is needed before assuming that a residual effect can be ignored. In general, it is done only when experience suggests that the effects of such treatments pass within a certain period, e.g., a single growing season, and when an analysis of variance raises no serious doubts that anything unusual is happening.

In such cases the method illustrated in Fig. 1 can be indefinitely repeated. Thus in the trial set out in Fig. 5, each plot consisted of one

Cc*C*	Ff*F*	Da*A*	Ee*D*	Bb*E*	Ad*B*		Fd*E*	Ef*C*	Ce*B*	Dc*A*	Ab*D*	Ba*F*
Ae*F*	Ea*C*	Dd*D*	Fc*E*	Cb*A*	Bf*B*		Fb*C*	De*A*	Bc*B*	Ca*E*	Af*D*	Ed*F*
Fe*E*	Bd*C*	Aa*D*	Db*B*	Ec*F*	Cf*A*		Cd*A*	Eb*F*	Df*E*	Ac*D*	Fa*B*	Be*C*

FIG. 5. EXAMPLE OF REPEATED CHANGES OF TREATMENTS USING ORTHOGONAL DESIGNS.

Key A-F 1951 Treatments. No residual effect remaining in 1953.
a-f 1952 Treatments. *A-F* 1953 Treatments.

apple tree of Cox's Orange Pippin, one of Worcester Pearmain and one of Laxton's Superb. Each year six spraying treatments were applied in order to observe and measure any injury they might cause. In the event of such an effect arising, it would be necessary to consider if the damage were permanent before proceeding further, but the design allows each year for a residual effect from the year before.

Non-orthogonal designs. A design similar except for the use of balance is exemplified by the spraying trial for the control of Fruit Tree Red Spider Mite, *Metatetranychus ulmi* (Koch), on apple trees, carried out at Writtle, Essex, England, and illustrated in Fig. 6. It has been described briefly before (Pearce, Jolly and Freeman, 1953). With this species of mite there may well be a residual effect of control measures in the following year because of the differential egg population left behind for the winter, but this is likely to disappear during the next summer when the mites disperse. Consequently each year's experiments should take into account possible residual effects from the treatments of the preceding year, but otherwise earlier treatments can be ignored. Of course, before assuming in any one season that an effect had in fact disappeared, it would be wise to make sure by carrying out an investigation on the mite population at the end of the previous season.

In this particular trial the first experiment, which took place in 1951, had seven treatments and was in Balanced Incomplete Blocks (Type T) as shown in Fig. 6, Diag. A. In the following year there were four treatments, the design being of Type T:OO as shown in Diag. B1; but it could equally have been of Type T:TT with seven treatments as shown in Diag. B2. In the third year the first set of treatments were plainly having no further effect. Consequently the design in Diag. B1 had degenerated to one of Type O. This was in fact changed to one of Type O:OT with four treatments as shown in Diag. C1, though it could have become one of Type O:TO as in Diag. C2. Equally, had a design of Type T:TT been adopted in 1952, it would have degenerated to one of Type T and could have been built up again as in the original design, Diags. C3 and C4, which correspond respectively to Diags. B1 and B2, showing the method.

As to efficiency factors, these are:

Type T	7/8
Type T:OO	1
Type T:TT	7/10 or 7/9*
Type O:OT	48/49
Type O:TO	7/8

*according to comparison.

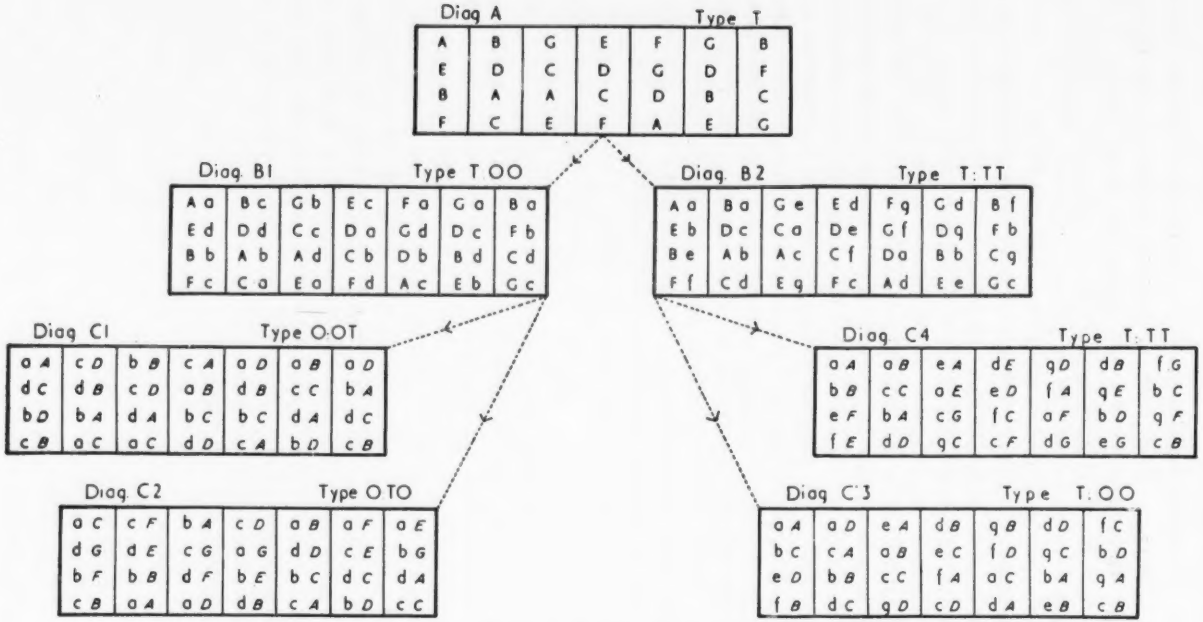

Diag A
Type T
A B G E F G B
E D C D G D F
B A A C D B C
F C E F A E G
Diag B1
Type T:OO
Aa Bc Gb Ec Fa Ga Ba
Ed Dd Cc Da Gd Dc Fb
Bb Ab Ad Cb Db Bd Cd
Fc Ca Ea Fd Ac Eb Gc
Diag B2
Type T:TT
Aa Ba Ge Ed Fg Gd Bf
Eb Dc Ca De Gf Dg Fb
Be Ab Ac Cf Da Bb Cg
Ff Cd Eg Fc Ad Ee Gc
Diag C1
Type O:OT
aA cD bB cA aD aB aD
dC dB cD aB dB cC bA
bD bA dA bC bC dA dC
cB aC aC dD cA bD cB
Diag C4
Type T:TT
aA aB eA dE gD dB fG
bB cC aE eD fA gE bC
eF bA cG fC aF bD gF
fE dD gC cF dG eG cB
Diag C2
Type O:TO
aC cF bA cD aB aF aE
dG dE cG aG dD cE bG
bF bB dF bE bC dC dA
cB aA aD dB cA bD cC
Diag C3
Type T:OO
aA aD eA dB gB dD fC
bC cA aB eC fD gC bD
eD bB cC fA aC bA gA
fB dC gD cD dA eB cB

FIG. 6. EXAMPLE OF REPEATED CHANGES OF TREATMENTS USING NON-ORTHOGONAL DESIGNS.
Key A–G 1951 Treatments. No residual effects remaining in 1953. *a* etc. 1952 Treatments. *A* etc. 1953 Treatments.

In general it is unwise to design a trial for numbers of treatments as diverse as four and seven, but in a trial like this the differences in replication can to a large extent be overcome by changing the intensity of sampling for the species under study. The important thing, however, is that the trial once started can be continued indefinitely, a new set of treatments being possible as soon as the residual effects of the last set but one have disappeared.

SUMMARY

Part I

(a) The design of new experiments with fruit trees is discussed and some general considerations advanced.

(b) The circumstances in which subsequent trials are possible are discussed.

(c) Possible designs are set out for a subsequent trial on the same material where the initial design is in randomized blocks.

(d) Attention is given to the designing of initial trials, so as to permit most readily of the addition of further sets of treatments.

Part II

Examples are given of some of the methods set out in Part I, using orthogonal and non-orthogonal designs both for the single changing of treatments and for repeated changing.

REFERENCES

Bose, R. C. and Nair, K. R. (1939). Partially balanced incomplete block designs *Sankhyā*, *4*, 337–72.

Pearce, S. C. (1952). Some new designs of Latin square type. *J. Roy. Statist. Soc.*, Ser. B, *14*, 101–6.

Pearce, S. C., Jolly, G. M. and Freeman, G. H. (1953). A review of experimental design at East Malling, 1948–1952. *Rep. E. Malling Res. Sta. for 1952*, 83–7.

Pearce, S. C. and Taylor, J. (1948). The changing of treatments in a long-term trial. *J. Agric. Sci.*, *38*, 402–10.

Pearce, S. C. and Taylor, J. (1950). The purposes and design of calibration trials. *Rep. E. Malling Res. Sta. for 1949*, 83–90.

Shrikhande, S. S. (1951). Designs for two-way elimination of heterogeneity. *Ann. Math. Statist.*, *22*, 235–47.

Yates, F. (1936a). Incomplete randomized blocks. *Ann. Eugen., Lond.*, *7*, 121–40

Yates, F. (1936b). Incomplete Latin squares. *J. Agric. Sci.*, *26*, 301–15.

VARIANCE HETEROGENEITY IN A RANDOMIZED BLOCK DESIGN

FRANKLIN GRAYBILL

Oklahoma A. and M. College

1. *Introduction.* The purpose of this paper is to give an exact criterion for testing whether all treatments are equal by a randomized block design under the follwing oconditions:

1. The error variances differ f om treatment to treatment.
2. The errors are correlated within a block but are independent from block to block.

In a randomized block experiment when we are confronted with the problem of heterogeneous error variances or correlated errors, the ratio of the mean square for the treatments to the mean square for the error (obtained by pooling the sum of squares for error) is not distributed as Snedecor's F. Hence this cannot be used as an exact criterion to test whether all the treatment means are equal. If all the assumptions for the analysis of variance (see (1)) are satisfied except the assumption of homogeneous error variance or the assumption of the errors being uncorrelated within a block, then we can still make certain exact tests. For example, any treatment comparison can be tested by sub-dividing the error sum of squares corresponding to that comparison and using Snedecor's F (see (2), (3), (4)).

However, the treatments are sometimes such that it is desired to test whether all the treatment means are equal rather than to test whether any specific treatment comparison is equal to zero. This was the case in a wheat varietal test conducted by the Oklahoma Agricultural Experiment Station[1].

Four varieties of wheat[2] were grown at each of thirteen locations

[1]I wish to thank Roy M. Oswalt, Assistant Agronomist, Oklahoma Agricultural Experiment Station, and A. M. Schlehuber, Senior Agronomist, Division of Cereal Crops and Diseases, and Oklahoma Agricultural Experiment Station for permission to use these data.

[2]This is a portion of more extensive data obtained as a part of the Oklahoma State-wide Small Grain Testing Program. There were some missing plots when the data were first collected. Therefore, some of the numbers have been filled in and some have been changed for purposes of illustration in this paper.

throughout the state and the yields in bushels per acre recorded. The yields are given in Table I with varieties designated simply as 1, 2, 3, and 4.

TABLE I

Location	Variety			
	1	2	3	4
1	43.60	24.05	19.47	19.41
2	40.40	21.76	16.61	23.84
3	18.08	14.19	16.69	16.08
4	19.57	18.61	17.78	18.29
5	45.20	29.33	20.19	30.08
6	25.87	25.60	23.31	27.04
7	55.20	38.77	21.15	39.95
8	55.32	34.19	18.56	25.12
9	19.79	21.65	23.31	22.45
10	46.24	31.52	22.48	29.28
11	14.88	15.68	19.79	22.56
12	7.52	4.69	20.53	22.08
13	41.17	32.59	29.25	43.95
Sums	432.84	312.63	269.12	340.13

To test the hypothesis that all variety means are equal, the location-variety means can be considered as a randomized block classification given by the mathematical model

$$(1.1) \qquad z_{ij} = \alpha + L_i + V_j + e_{ij} \; ; \; j = 1, 2, 3, 4; \; i = 1, 2, \cdots, 13;$$

where z_{ij} is the yield of variety j at location i, α is the general mean, L_i is the effect of the ith location (assumed random), V_j is the effect of the jth variety, and the e_{ij} are the error terms which are assumed to be normally distributed with means zero, variances σ_j^2 , and such that the correlation between e_{ij} and e_{km} is zero if $i \neq k$.

Disregarding the fact of heterogeneous error variances and correlated errors, the conventional analysis of variance for testing whether all variety means are equal yields an $F = 6.56$.

Although this F value is significant at the 1 per cent level, it appears from the data in Table I that the error variance is not homogeneous. This is clear from the fact that the estimates of σ_1^2 , σ_2^2 , σ_3^2 , and σ_4^2 are 145.86, -14.18, 75.29, and 18.25, respectively.

The effect, in general, of ignoring heterogeneity of error variances in the usual analysis of variance is that the tabulated F is too low and hence we get significance more frequently than we should. For example, our observed F is 6.56 and the tabulated F at the 1 per cent level is 4.38 for 3 and 36 d.f. Thus we do not know whether there is really significance at the 1 per cent level or whether the large F value is due to the heterogeneous error variances. Since in this example the observed F is 50 per cent greater than the tabulated F, we would probably be willing to conclude that there are variety differences; but if, for example, the observed F had been between the 1 per cent and 5 per cent value of the tabulated F and since error variances are heterogeneous, then it would be difficult to draw conclusions about variety differences.

2. *Theory.* Let X_i be a $(p + 1) \times 1$ column vector with elements $x_{ij}(i = 1, 2, \cdots, n; j = 1, 2, \cdots, p + 1)$ such that X_i has a multivariate normal distribution with covariance matrix $\sum_1$, and with a mean equal to the $(p + 1) \times 1$ vector U_z with elements $\mu_t(t = 1, 2, \cdots, p + 1)$. Also we will assume the X_i and X_k are independent if $i \neq k$; that is to say $X_i(i = 1, 2, \cdots, n)$ is a random sample of size n from the above specified $(p + 1)$ variate normal distribution.

P. L. Hsu (5) has shown that the hypothesis $\mu_1 = \mu_2 = \cdots = \mu_{p+1}$ can be tested in the following manner. Define a new $p \times 1$ vector Y_i with elements y_{ij} such that

$$y_{ij} = x_{ij} - x_{i,p+1} \qquad j = 1, 2, \cdots, p; i = 1, 2, \cdots, n. \tag{2.1}$$

Then Y_i can be considered a random sample of size n from a normal distribution with covariance matrix $\sum_2$, and with mean equal to the $p \times 1$ vector U_Y with element $\mu_t - \mu_{p+1}(t = 1, 2, \cdots, p)$. To test $\mu_1 = \mu_2 = \cdots = \mu_{p+1}$, we use the transformed variables Y_i and test the hypothesis U_Y equal to the zero vector by Hotelling's T^2 test (6). That is to say, under the hypothesis $\mu_1 = \mu_2 = \cdots = \mu_{p+1}$ (and if $n > p$) then the quantity

$$\overline{Y}'\left[\sum_{i=1}^{n}(Y_i - \overline{Y})(Y_i - \overline{Y})'\right]^{-1}\overline{Y}\,\frac{(n-p)n}{p} = F \tag{2.2}$$

has Snedecor's F distribution with p and $n - p$ degrees of freedom where $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, where $\overline{Y}'$ indicates the transpose of the matrix $\overline{Y}$, and where a matrix with an exponent of minus 1 signifies the inverse of the matrix.

To illustrate how Hotelling's T^2 can be used to test $V_1 = V_2 = V_3 = V_4$ in the model given by (1.1) where the error variance differs from variety to variety, put the z_{ij} of (1.1) equal to the x_{ij} of (2.1) where $p = 3$ and $n = 13$.

This gives

$$y_{ij} = z_{ij} - z_{i4} = V_j - V_4 + e_{ij} - e_{i4} \tag{2.3}$$

and the matrix U_Y has for its elements $V_j - V_4$. Thus testing U_Y equal to the zero vector is equivalent to testing $V_j = V_4 (j = 1, 2, 3)$, i.e. to testing all V_i equal.

Since z_{ij} in Table I represents the average yield of variety j at location i, we use (2.3) to compute the elements y_{ij} which are tabulated in the ith row and jth column of (2.4).

$$\begin{matrix} 24.19 & 4.64 & 0.06 \\ 16.56 & -2.08 & -7.23 \\ 2.00 & -1.89 & 0.61 \\ 1.28 & 0.32 & -0.51 \\ 15.12 & -0.75 & -9.89 \\ -1.17 & -1.44 & -3.73 \\ 15.25 & -1.18 & -18.80 \\ 30.20 & 9.07 & -6.56 \\ -2.66 & -0.80 & 0.86 \\ 16.96 & 2.24 & -6.80 \\ -7.68 & -6.88 & -2.77 \\ -14.56 & -17.39 & -1.55 \\ -2.78 & -11.36 & -14.70 \end{matrix} \tag{2.4}$$

The ith row of (2.4) is the transpose of the vector Y_i. Thus we get

$$\overline{Y}' = (7.13 \quad -2.12 \quad -5.46). \tag{2.5}$$

We also compute the matrix

$$\sum_{i=1}^{13} (Y_i - \overline{Y})(Y_i - \overline{Y})' = \begin{bmatrix} 2{,}151.8706 & 894.5435 & -274.1924 \\ 894.5435 & 542.1045 & 36.3449 \\ -274.1924 & 36.3449 & 446.3738 \end{bmatrix}$$

and the inverse

$$(2.6) \qquad \begin{bmatrix} .0023572760 & -.0040087773 & .0017744001 \\ -.0040087773 & .0086721040 & -.0031685618 \\ .0017744001 & -.0031685618 & .0035882215 \end{bmatrix}$$

Substituting (2.5) and (2.6) into (2.2) and evaluating, we get

$$F = 7.59$$

with associated degrees of freedom of 3 and 10, which is larger than Snedecor's tabulated F at the one per cent level. Thus we conclude that the large F is due to the variety means being different. This is an exact criterion for testing the equality of means in a randomized block classification when the error variances differ from treatment to treatment and when the errors may be correlated within a block but are uncorrelated from block to block.

It should be emphasized that this test is valid only if $n > p$. It is immaterial which element is subtracted in (2.1); that is to say the test is invariant if $x_{i,p+1}$ in (2.1) is replaced by any x_{it} ($t = 1, 2, \cdots, p + 1$; $t \neq j$).

REFERENCES

1. C. Eisenhart, "The Assumptions Underlying the Analysis of Variance," *Biometrics*, *3*, 1–21, 1947.
2. W. G. Cochran and G. M. Cox, *Experimental Designs*, John Wiley & Sons, 71–72, 1950.
3. O. Kempthorne, *The Design and Analysis of Experiments*, John Wiley & Sons, 176–177, 1952.
4. G. W. Snedecor, *Statistical Methods*, The Iowa State College Press, 412–413, 1946.
5. P. L. Hsu, "Notes on Hotelling's Generalized T," *Annals of Mathematical Statistics*, *9*, 241–242 (c), 1938.
6. H. Hotelling, "The Generalization of Student's Ratio," *Annals of Mathematical Statistics*, *2*, 359–378, 1931.

SOME OBSERVATIONS ON SENSITIVITY AND REPEATABILITY OF TRIAD TASTE DIFFERENCE TESTS[1]

J. W. HOPKINS

Division of Applied Biology, National Research Council of Canada

INTRODUCTION

The literature on taste difference experimentation is now extensive (3) and statistical testing of null hypotheses of "no discrimination" well understood. Publications dealing quantitatively with the sensitivity and repeatability of such experiments, and with effects of subject training and experience on these characteristics, are however less numerous. Some results obtained in the writer's laboratory from 2,304 unit observations in tests of potential subjects for sensory difference appraisals of processed whole milks are accordingly reported below.

EXPERIMENTAL

Experimental Taste Contrasts

Two experiments were made. Both utilized sensory difference appraisals of the "triangular" variety, in which subjects attempt to identify the like and unlike items in coded triads comprising 2 aliquots of one test material and 1 of another.

Experiment I involved 8 contrasts between pasteurized homogenized fresh whole milk with and without a single primary taste additive in one of the following gram-molecular concentrations:

Bitter (Bi): 0 vs. (1) .000002M or (2) .000004M quinine sulfate
Salt (Sa): 0 vs. (1) .005M or (2) .010M sodium chloride
Sour (So): 0 vs. (1) .005M or (2) .010M lactic acid
Sweet(Sw): 0 vs. (1) .025M or (2) .050M lactose

[1]N. R. C. Publication No. 3432

In Experiment II all test aliquots of milk contained a primary taste additive, in concentrations adjusted in geometric progression to give 8 additional contrasts as follows:

Bi: (1) .0000020M or (2) .0000049M vs. (3) .0000120M quinine sulfate
Sa: (1) .005M or (2) .010M vs. (3) .020M sodium chloride
So: (1) .0050M or (2) .0087M vs. (3) .0150M lactic acid
Sw: (1) .025M or (2) .047M vs. (3) .088M lactose

Method of Tasting

Before presentation for tasting, all aliquots were brought to 60°F. Tasting and recording of results was done independently by each subject in a laboratory maintained at an air temperature of 76°F. (dry bulb) and a relative humidity of 52%.

Twelve subjects participated, 3 men and 3 women in Experiment I, and 4 men and 2 women in Experiment II. All were scientists or technicians without previous experience in sensory difference testing. They appraised the 8 test contrasts in their experiment on each of 24 days, 4 appraisals being made at a mid-morning and 4 at a mid-afternoon session. At each session, "warm-up" tasting of contrasts involving the maximal experimental concentration of each additive immediately preceded appraisal of the test triads. For the first 12 days of both experiments this "warm-up" was unsupervised, instruction being confined to an initial outline of the elements of tasting technique. On the following 12 days, however, it was used for supervised practice and specific training in detection of each taste difference.

Sequence of Tasting

In both experiments, each individual's tasting schedule was independently randomized subject to three balancing conditions. (i) All four additives were tasted at each session, once in each permutation of the sequence Bi, Sa, So, Sw in each half of the experiment. (ii) Morning and afternoon sessions of the same day were complementary half-replicates of the eight taste contrasts, and the six combinations of kind and quantity of additive providing half-replicates satisfying the preceding condition each occurred once in every quarter of the experimental period. (iii) Each of the aliquot triad sequences aab, aba, baa, abb, bab, bba providing each flavor contrast (a) v. (b) was also presented once in each quarter. A typical resulting daily schedule (that for

Subject C on Day 5) was:

morning:	1st triad	Bi	(2)	(0)	(0)
	2nd triad	Sw	(1)	(1)	(0)
	3rd triad	So	(1)	(0)	(1)
	4th triad	Sa	(0)	(0)	(2)
afternoon:	1st triad	Sa	(1)	(1)	(0)
	2nd triad	Sw	(2)	(0)	(0)
	3rd triad	So	(2)	(0)	(0)
	4th triad	Bi	(0)	(1)	(0)

STATISTICAL ANALYSIS OF RESULTS

Experiment I

Sensory Detection of Additives

In the aggregate of 144 appraisals of each contrast (24 by each subject), duplicate aliquots were matched with the frequencies listed in the fourth and seventh columns of Table I. These recorded frequencies were clearly inconsistent with purely random guessing, which would require them to be binomial variates with expectation $144(0.3\dot{3}) = 48$ and standard deviation $[144(0.3\dot{3})(0.6\dot{6})]^{1/2} = 5.66$. Their smallest attained this expectation, and all others exceeded it by more than 3 times the specified s.d. Also, uniformly more matchings were recorded for contrasts (2) v. (0) than for (1) v. (0).

Effect of Training

Aggregate frequencies of matching recorded in all 72 appraisals of each contrast (12 by each subject) during Days 1–12 and 13–24 respectively are also listed in Table I. These were strongly suggestive of

TABLE I

AGGREGATE FREQUENCY OF MATCHING DUPLICATE ALIQUOTS IN 144 TRIAD APPRAISALS, EXPERIMENT I

Taste additive	Contrast (1, 0)			Contrast (2, 0)		
	Days 1–12	Days 13–24	Days 1–24	Days 1–12	Days 13–24	Days 1–24
Bi	23	25	48	29	44	73
Sa	35	41	76	49	58	107
So	25	41	66	40	49	89
Sw	33	41	74	48	64	112

greater discrimination in the second period. For all eight contrasts the second frequency exceeded the first, whereas with identity of corresponding probabilities of matching, the chance of fortuitous occurrence of eight differences of like sign would be only 1 in 2^7.

Repeatability of Judgments

If the probability p_{ijk} of matching duplicate aliquots in a single appraisal of the ith contrast by the jth subject on the kth day, although in general unequal for different j, was constant for specified i and j over a period $k = r, \cdots, s$, the daily totals $T_{i.k}$ of matchings recorded during this period would for specified i be binomial variates of Poisson (1, sec. 23). Cochran (2) has shown that in these circumstances the index of dispersion $Q_i = c(c-1)\sum_{k=r}^{s}(T_{i.k} - T_{i..}/c)^2/(cT_{i..} - \sum_j T_{ij.}^2)$, in which $c = s - r + 1$, will have a random sampling distribution approximately close to that of χ^2 with $s - r$ d.f.

In the present instance, the eight Q_i for Days 1–12 were numerically consistent, both individually and collectively, with purely random fluctuation in the $T_{i.k}$. Six of the eight Q_i for Days 13–24 however exceeded the 26% point of tabular χ^2 for 11 d.f. (5). The probability of so disproportionate a distribution of Q_i from statistically stable $T_{i.k}$ would be less than 0.005 (7). A plot of the daily totals $T_{..k}$ for

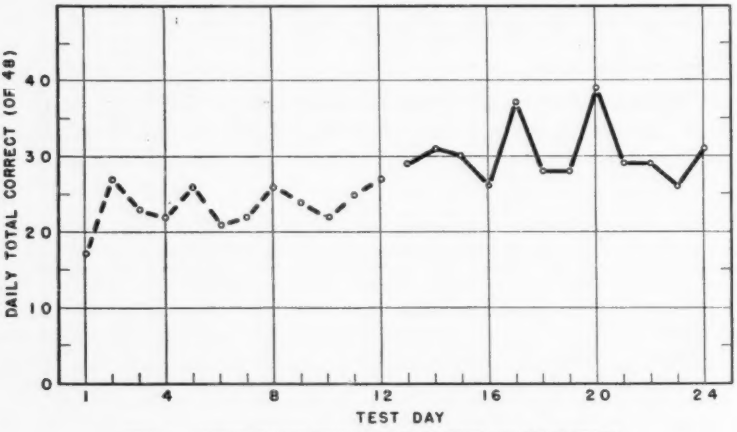


FIG. 1. AGGREGATE DAILY FREQUENCY OF MATCHING

all i and j (Fig. 1) shows notably aberrant results for Days 17 and 20, but no consistent trend suggestive of a cumulative effect of training or habituation within either period.

Differential Acuity

Table II lists the frequency of matching recorded for each subject in all appraisals of Bi, Sa, So and Sw contrasts during each half of the

TABLE II

FREQUENCY OF MATCHING IN 24 APPRAISALS BY SUBJECT, PERIOD AND TASTE CONTRAST, EXPERIMENT I

Period and Contrast	Subject					
	A	*B*	*C*	*D*	*E*	*F*
Days 1–12:						
Bi (1, 0) + (2, 0)	9	5	8	18	6	6
Sa (1, 0) + (2, 0)	18	19	11	13	13	10
So (1, 0) + (2, 0)	13	11	5	18	14	4
Sw (1, 0) + (2, 0)	18	9	14	15	15	10
Days 13–24:						
Bi (1, 0) + (2, 0)	17	6	11	15	13	7
Sa (1, 0) + (2, 0)	22	18	19	16	11	13
So (1, 0) + (2, 0)	15	8	12	24	20	11
Sw (1, 0) + (2, 0)	23	12	13	20	20	17

experiment. These were suggestive of marked individual differences in sensitivity to specific contrasts, not eliminated by the training given during Days 13–24; also of some degree of correlation in individual sensitivity to all contrasts, in both periods. As however the listed frequencies are binomial variates of Poisson of unknown but clearly unequal variance, statistical testing of these indications was restricted to the following analysis of the Fisher-Yates (4) normalized rank scores for the six items in each row.

Denote by X_{ij} the resulting score for the frequency in the ith row and jth column of the table, and let

$$X_{ij} + X_{(i+4),j} = T_{ij} \qquad (i = 1, \cdots 4)$$

$$\sum_{i=1}^{4} X_{ij} = U_{\cdot j} \qquad \sum_{i=5}^{8} X_{ij} = V_{\cdot j}$$

Then (a) on the null assumption of an independent subject-ranking for the same primary taste after and before training, the six T_{ij} for specified $i = 1, \cdots 4$ must be random variates with expectations uniformly zero and the between-j: remainder variance ratio of the

corresponding X_{ij} , with 5:5 d.f., thus provides a test of independence. Alternatively the indications (sums of squares) for all four tastes may be combined in a single variance ratio with 20:20 d.f. (b) On the null hypothesis of an independent subject-ranking for each of the four tastes within periods, the $U_{.j}$ and $V_{.j}$ must likewise be random variates with expectations uniformly zero, and the between-j: remainder variance ratio of the corresponding X_{ij} with 5:15 d.f. thus provides a test of the actual randomness of either the U or V. In this instance test (a), second alternative, gives a variance ratio of 6.47, exceeding the 0.1% point of tabular $F_{20:20}$. Test (b) gives variance ratios of 3.20 and 4.21 respectively for Days 1–12 and for Days 13–24. These exceed the 5% and 2.5% points of tabular $F_{5:15}$.

Subjective Patterns

Total frequencies of matching recorded for each subject in an aggregate of 32 appraisals with each type of triad are listed in Table III. Here again, had all corresponding probabilities been identical for each

TABLE III

AGGREGATE FREQUENCIES OF MATCHING DUPLICATE ALIQUOTS WITH (a) AND WITHOUT (b) TASTE ADDITIVES IN VARIOUS TYPES OF TRIAD, EXPERIMENT I

Subject	Type of triad						Index of discrepancy
	baa	*aba*	*aab*	*bba*	*bab*	*abb*	*Q*
A	21	23	24	23	22	22	1.0
B	15	13	19	11	13	17	6.2
C	13	14	10	21	16	19	11.0
D	23	23	21	23	28	21	8.4
E	18	14	18	20	19	23	6.1
F	11	14	17	14	6	16	11.4

type, the frequencies in each row of the table would have been binomial variates of Poisson. In fact two of the six Cochran indices of dispersion Q for these row frequencies tabulated in the last column exceed the 5.2% point of tabular χ^2 for 5 d.f., an outcome having a chance probability of less than 0.04 for homogeneous data. Subject C (female) was consistently more sensitive to triads comprising 2 b and 1 a aliquot (b = no additive) than to those comprising 2 a and 1 b, while F (male) was noticeably resistant to the sequence *bab*.

Experiment II

Detection of Differences

Frequencies of matching recorded in the aggregate of 144 appraisals of each taste contrast are reported in the fourth and seventh columns of Table IV. The smallest of these, namely 85 for Bi (3, 2), exceeds

TABLE IV

AGGREGATE FREQUENCY OF MATCHING DUPLICATE ALIQUOTS IN 144 TRIAD APPRAISALS, EXPERIMENT II

Taste additive	Contrast (3, 2)			Contrast (3, 1)		
	Days 25–36	Days 37–48	Days 25–48	Days 25–36	Days 37–48	Days 25–48
Bi	40	45	85	45	63	108
Sa	40	57	97	49	63	112
So	38	50	88	43	55	98
Sw	38	52	90	63	68	131

the no-discrimination expectation of 48 by some 6 times the corresponding standard deviation. For all four tastes, the frequency of matching was greater for the contrast (3) v. (1) than for (3) v. (2).

Effect of Training

Table IV also shows that as in the preceding experiment, the frequency of matching recorded for each of the eight taste contrasts was greater in the second 72 appraisals with specific training (Days 37–48) than in the first 72 without (Days 25–36).

Repeatability of Judgments

Indices Q_i of dispersion in the daily total frequencies of matching, calculated as before for each of the eight contrasts i and for Days 25–36 and 37–48 separately, were individually and collectively consistent with sensibly constant p_{ijk} within both periods. The largest of these 16 Q_i , namely 18.8, fell below the 6% point of χ^2 for 11 d.f., while their total $\sum Q_i = 153.7$ was less than the expectation of χ^2 for 176 d.f.

Differential Acuity

Table V is indicative of individual differences similar to those noted in Table II. Similar rank analyses gave variance ratio criteria of 2.86

TABLE V

FREQUENCY OF MATCHING IN 24 APPRAISALS BY SUBJECT, PERIOD AND TASTE CONTRAST, EXPERIMENT II

Period and Contrast	Subject					
	G	*H*	*I*	*J*	*K*	*L*
Days 25–36:						
Bi (3, 2) + (3, 1)	23	14	13	8	12	15
Sa (3, 2) + (3, 1)	17	15	21	13	12	11
So (3, 2) + (3, 1)	23	7	18	14	9	10
Sw (3, 2) + (3, 1)	21	17	19	16	14	14
Days 37–48:						
Bi (3, 2) + (3, 1)	19	19	21	23	8	18
Sa (3, 2) + (3, 1)	23	19	22	17	21	18
So (3, 2) + (3, 1)	21	17	19	23	7	18
Sw (3, 2) + (3, 1)	24	19	20	22	16	19

(20:20 d.f.) for differences between subjects in respect of specific tastes common to both periods, and of 5.33 and 3.43 (5:15 d.f.) for differences common to all four tastes in the first and in the second half of the experiment, which respectively exceeded the 2%, 1% and 5% points of corresponding tabular F.

Subjective Patterns

Total frequencies of matching recorded by each subject for each type of triad are enumerated in Table VI. Also listed in this table are

TABLE VI

FREQUENCY OF MATCHING DUPLICATE ALIQUOTS IN 32 APPRAISALS OF TRIADS COMPRISING CONCENTRATIONS a AND b ($a > b$) OF FLAVOR ADDITIVES EXPERIMENT II

Subject	Type of triad						Index of discrepancy Q
	baa	*aba*	*aab*	*bba*	*bab*	*abb*	
G	29	28	29	26	28	31	4.6
H	23	23	22	24	16	19	7.2
I	29	21	28	28	23	24	11.6
J	19	23	19	25	25	25	8.7
K	19	16	20	12	16	16	5.8
L	17	16	15	24	25	26	18.6

the appropriate indices of dispersion Q calculated as for Table III. Again two of the latter exceeded the 5% point, 11.1, of tabular χ^2 for 5 d.f. Like *C* in Experiment I, *L* (male) was consistently more sensitive to triads comprising 1 *a* and 2 *b* aliquots (concentration of taste additive $a > b$) than to those comprising 2 *b* and 1 *a*. Subject I (male) was resistant to the sequence *bab*, thus resembling *F*, and also to *aba*.

DISCUSSION

Despite their subjective element, the judgment-reactions recorded in the first half of Experiment I and in both halves of Experiment II seem to have been sensibly statistically stable. The anomalies on Days 17 and 20 of the first experiment may therefore be plausibly attributed to technical factors, e.g. peculiarities of the milk used on these occasions.

It is noteworthy that no effect of increasing familiarity with the taste contrasts *per se* on group acuity was demonstrable in these trials. In respect of the training described, it may be noted that the probability p_{ijk} of matching duplicate aliquots in a single triad was the resultant

$$p_{ijk} = p_{1ijk} + p_{0ijk}(1 - p_{1ijk})$$

of the probabilities p_{1ijk} of actual sensory discrimination and p_{0ijk} of correct guessing in the absence of discrimination, here uniformly 1/3. The over-all aggregates of 282 and 363 matchings in 576 appraisals recorded in the first and second halves of Experiment I accordingly yield estimates $\hat{p}_{1\ldots}$ of 0.235 and 0.445 for $p_{1\ldots}$, the mean of p_{1ijk}, for $k = 1, \cdots, 12$ and $k = 13, \cdots, 24$ respectively. For the first and second halves of Experiment II, the over-all frequencies of matching were 356 and 453, giving $\hat{p}_{1\ldots} = 0.427$ and 0.679. The training given may thus have increased the actual probabilities $p_{1\ldots}$ characterizing group mean acuity by a factor of $1\frac{1}{2}$ or 2.

Persistence of individual differences in sensitivity after training may also be remarked upon. The maximal and minimal individual aggregate frequencies of matching during the second period of Experiment I were 77/96 by *A* and 44/96 by *B*, and during the second period of Experiment II 87/96 by *G* and 52/96 by *K*. Had all recorded frequencies been equal to their expectations, these would have corresponded to mean $P_{1ij.}$ of 0.70 and 0.19 for *A* and *B* respectively, and of 0.86 and 0.31 for *G* and *K*. Some allowance must be made for exaggeration in the indicated ratios of $3\frac{1}{2}$:1 for *A* and *B* and of $2\frac{1}{2}$:1 for *G* and *K*, due to selection of the largest and smallest recorded frequencies in each group of 6. Nevertheless, these results are suggestive of individual

differences of the same order of magnitude as those between group averages before and after training.

Helm and Trolle (6) observed a differential reaction to triads comprising 1 *a* and 2 *b* and 2 *a* and 1 *b* aliquots in experimental beer tasting. They did not however distinguish between individuals in this regard, nor did they observe the resistance to *aba* or *bab* sequences noted here for Subjects *F* and *I*.

ACKNOWLEDGMENTS

Elinor M. Zuckerman supervised the technical details of both experiments. The writer is also indebted to the 12 participants for patient and conscientious collaboration.

REFERENCES

1. Aitken, A. C. *Statistical Mathematics*, 4th ed., Oliver and Boyd, Edinburgh and London, 1945.
2. Cochran, W. G. The comparison of percentage in matched samples. *Biometrika, 37:* 256–266, 1950.
3. Dawson, Elsie, H. and Harris, Betsy, L. Sensory methods for measuring food quality; review of literature and proceedings of conference. *Agric. Information Bull. 34 (iv)*, U. S. Dept. of Agriculture, Washington, D. C., 1951.
4. Fisher, R. A. and Yates, F. *Statistical Tables for Biological, Agricultural and Medical Research, 3rd ed.*, Oliver and Boyd, Edinburgh and London, 1948.
5. Hartley, H. O. and Pearson, E. S. Tables of the χ^2-integral and of the cumulative Poisson distribution. *Biometrika, 37:* 313–325, 1950.
6. Helm, E. and Trolle, B. Selection of a taste panel. *Wallerstein Lab. Comm. 9:* 181–194, 1946.
7. National Bureau of Standards. *Applied Mathematics Series 6, Tables of the Binomial Probability Distribution.* U. S. Government Printing Office, Washington, D. C., 1950.

THE ESTIMATION OF RESPONSE-TIME DISTRIBUTION III TRUNCATION AND SURVIVAL

M. R. Sampford

*Agricultural Research Council Unit of Statistics, University of Aberdeen**

1. INTRODUCTION

A general survey of the problems involved in the estimation of response-time distributions was given in the first paper of this series (Sampford, 1952a). Many of these problems arise as consequences of failure to observe the desired response in all subjects. The courses to be followed when this failure results from the operation of a stimulus other than that under investigation—when, for example, animals die from an infection, as well as from the application of a poison—have been discussed in a second paper (Sampford, 1952b). However, incomplete data may also occur as a result of *truncation*, when the experimenter fails, intentionally or accidentally, to observe the response on some individuals, or of *survival*, a convenient term adopted from the particular case in which the response is death, and indicating a failure on the part of some individuals to show the response, either because they are immune to the treatment or because they recover from its effects. These two problems are discussed in this paper.

In future references the two papers cited above will be denoted simply by (I) and (II). A detailed list of conventions, definitions, and symbols used consistently throughout the series was given in (II).

Throughout this paper the word 'time' will be used indifferently for true time and for a transformed time-metameter; the meaning in any instance should be clear from the context, or from the symbols used, τ denoting a measurement in 'true time', t some function of τ, normally distributed when the stimulus under investigation operates without interference. (Usually $t = \tau$, $\log \tau$, or $1/\tau$).

*Formerly of the Lectureship in the Design and Analysis of Scientific Experiment, University of Oxford.

2. THE PROBLEM OF TRUNCATION

2.1 *A survey of the literature*

Samples in which individual observations are available only over parts of the range of an observed variable are not uncommon. The situation in which individual values are available only at the two extremes of the range has some industrial applications; it is unlikely to occur, however, in response-time studies, where incomplete samples are almost always of the type in which individual observations are not available for one or both of the extremes. Halperin (1952) has discussed at length the theory of maximum likelihood estimation from such samples, Walsh (1950) has discussed non-parametric estimation, Bliss and Stevens (1937), Cohen (1950), Cohen and Woodward (1953), and Hald (1952, §6.9) have suggested methods for the solution of the maximum likelihood equations in the normal case, and Gupta (1952) has derived best linear estimates for some types of normal sample.

2.2 *The occurrence of truncated samples*

A series of response-time data may be truncated at the lower end of the time scale if a stimulus takes effect with unexpected rapidity, so that some individuals respond before detailed observation is begun. Truncation at the lower end of the scale can also occur as a result of transformation, as when response-times are transformed reciprocally, after truncation at the upper end of the time scale. Again, when a grouped distribution is transformed logarithmically, any observations in the group $0 < \tau \leq \tau_1$ will fall in an infinite negative tail of the distribution of log times. I have already suggested (I, §3.5) that an attempt to deal with such observations by attaching a class-mark to them would be unwise, and that the sample should be analysed as though truncated at the 'time' $t_1 = \log \tau_1$. Truncation at the upper end of the scale is usually deliberate. Economic considerations may necessitate the suspension of observations at a pre-determined time, or when a certain proportion of the sample members have shown the response; alternatively it may be desired to perform an interim analysis on an experiment still in progress, in which case the data may be considered as truncated at the time of analysis.

A theoretical distinction can be drawn between data truncated at a fixed time and at a fixed sample point (for example, after 75% of individuals have responded); the distinction has, however, little practical importance, the methods of analysis for the two situations being effectively identical.

A more important distinction lies between the case in which the numbers of observations falling in the truncated tails are known, and that in which they are unknown. It may sometimes happen in response-time-studies that the size of a truncated tail is unknown; this situation could arise, for example, in the study of time of appearance of tumours on mice treated with a carcinogenic material, the tumours themselves counting as individuals, or in studying the time required to capture the individuals of an isolated population of wild animals. In such cases, if observation is suspended before all individuals have been observed, or before all tumours have developed, the number remaining is unknown. (A further example is discussed in §3.2.) Such data can be analysed by methods given by, among others, Cohen (1950), Cohen and Woodward (1953), and Hald (1952), who restricts the description 'truncated' to samples of this type, suggesting the alternative term 'censored' for data with known numbers in the truncated tails. In response-time studies, however, the total number of individuals in the sample is usually known, and the present discussion will be largely restricted to this situation.

2.3 *The normal distribution: the Bliss-Stevens method*

Consider a sample truncated at 'times' $t = T_1$ and $t = T_2$, with n_1, n, and n_2 individuals in the three parts of the range, so that for any individual the available information can take one of three forms:

$$\begin{aligned} &t < T_1 && (n_1 \text{ individuals}) \\ &t = t_i\,; \quad T_1 \leq t_i \leq T_2\,; && i = 1 \cdots n \\ &t > T_2 && (n_2 \text{ individuals}), \end{aligned}$$

where in particular cases T_1 may become $= -\infty$ or $T_2 = +\infty$, with the corresponding $n = 0$. The distribution of t is supposed normal,

$$f(t)\,dt = \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{(t-\mu)^2}{2\sigma^2} = \frac{\beta}{\sqrt{2\pi}} \exp - \frac{1}{2}(\alpha + \beta t)^2,$$

and the parameters of the distribution are to be estimated from the truncated sample.

Bliss and Stevens (1937) use a maximum likelihood iterative procedure leading to the equations

$$\left.\begin{aligned} Ni_{\mu\mu}\,\delta m + Ni_{\mu\sigma}\,\delta s &= \frac{1}{s^2}\sum_1^n (t_i - m) - \frac{n_1 Z_1}{sP_1} + \frac{n_2 Z_2}{sQ_2} \\ Ni_{\mu\sigma}\,\delta m + Ni_{\sigma\sigma}\,\delta s &= \frac{1}{s^3}\sum_1^n (t_i - m)^2 - \frac{n}{s} - \frac{n_1 Z_1 y_1}{sP_1} + \frac{n_2 Z_2 y_2}{sQ_2} \end{aligned}\right\} \quad (1)$$

where

$$N = n_1 + n + n_2 ,$$

$$y_i = (T_i - m)/s$$

$$Z_i = \frac{1}{\sqrt{2\pi}} e^{-y_i^2/2}, \qquad P_i = 1 - Q_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_i} e^{-u^2/2}\, du \qquad (2)$$

the i's are certain functions of s and of y_1 and y_2 and the corresponding Z's, P's, and Q's, and δm, δs are increments to be added to provisional estimates m, s of μ and σ. The asymptotic variance-covariance matrix of the estimates is obtained as the inverse of the coefficient matrix calculated in the final cycle. The i's can be calculated directly from tables of Z and P, but the labour can be considerably shortened by expressing them either in terms of the incomplete normal moment function (Pearson, 1930, Table IX), the form in which they are given by Bliss and Stevens, or of the functions λ and ζ, defined and tabulated by the present author (**II**), when they are given by

$$\left.\begin{aligned} s^2 i_{\mu\mu} &= (1 - P_1 - Q_2) + P_1\lambda(-y_1) + Q_2\lambda(y_2) \\ s^2 i_{\mu\sigma} &= P_1\zeta(-y_1) - Q_2\zeta(y_2) \\ s^2 i_{\sigma\sigma} &= 2(1 - P_1 - Q_2) + P_1 y_1\zeta(-y_1) - Q_2 y_2\zeta(y_2). \end{aligned}\right\} \qquad (3)$$

2.4 *A modification of the Bliss-Stevens method*

The coefficients of (1) are expected values. They may be replaced, if convenient, by values calculated in terms of the observed sample, without affecting the final estimates. Garwood (1941) has shown that, in the analysis of quantal response data by the probit method, this replacement has the effect of reducing the number of cycles required for convergence. It seems likely that this result is generally true, and that this modification is therefore desirable, provided that it does not noticeably increase the labour of a single cycle. In the present case, in fact, using the tables of λ and ζ mentioned above, the labour, even of a single cycle, is slightly reduced, provided the method is also modified to give immediate estimates of α and β, rather than of μ and σ. In this notation the logarithm of the likelihood is

$$\log L = C + n_1 \log P_1 + n \log \beta - \sum_1^n (\alpha + \beta t_i)^2 + n_2 \log Q_2 .$$

When $T_1 = -\infty$ and $n_1 = 0$, the situation reduces to a particular case of the 'accidental death' model considered in the second paper of this series, and has already been discussed in that connection (**II**, §5.3). By a manipulation similar to that employed there, the iterative equa-

tions for the general case can be reduced to the form

$$\left.\begin{aligned} D_1 a_2' &= n_1\zeta(-y_1) - n_2\zeta(y_2) \\ D_2 b_2 &= n_1(T_1 - \bar{t})\zeta(-y_1) - n_2(T_2 - \bar{t})\zeta(y_2) + (2n/b_1) \end{aligned}\right\} \quad (4)$$

where

$$D_1 = n_1\lambda(-y_1) + n + n_2\lambda(y_2)$$

$$D_2 = \left\{(n/b_1^2) + n_1T_1^2\lambda(-y_1) + \sum_1^{n} t_i^2 + n_2T_2^2\lambda(y_2)\right\} - D_1\bar{t}^2$$

$$\bar{t} = \left\{n_1T_1\lambda(-y_1) + \sum_1^{n} t_i + n_2T_2\lambda(y_2)\right\} \Big/ D_1$$

$$y_j = a_1 + b_1T_j\,, \qquad (j = 1, 2)$$

and a_1 and b_1 are provisional estimates of α and β,

$$a_2 = a_2' - b_2\bar{t}$$

and b_2 are the revised estimates, and

$$\lambda(\eta) = \nu^2(\eta) - \eta\nu(\eta)$$

$$\zeta(\eta) = \nu(\eta) - \eta\nu(\eta),$$

where

$$\nu(\eta) = Z(\eta)/Q(\eta).$$

The final estimates a' and b are uncorrelated, and their asymptotic variances are estimated by the reciprocals of their coefficients D_1 and D_2 in (4), as calculated in the final cycle.

2.5 *Other methods*

Hald (1952) considers only truncation at one tail, and obtains a single maximum likelihood equation in α, which may be solved with the aid of a function $\psi'(x)(= \nu(-x))$, (Hald, 1952b, Table X). A second equation is then explicitly soluble for σ, and tables are given from which the variances and covariance of σ and $\mu = -\alpha\sigma$ can be obtained.

Cohen considers truncation at both ends of the range, and obtains maximum likelihood equations for α and σ, which are solved by an iterative process different from those used above. The variances and covariances of the estimates are then calculated separately.

All four methods here described are fundamentally maximum likelihood methods, and will thus give identical estimates for a given

where

$$N = n_1 + n + n_2 ,$$

$$y_i = (T_i - m)/s$$

$$Z_i = \frac{1}{\sqrt{2\pi}} e^{-y_i^2/2}, \qquad P_i = 1 - Q_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_i} e^{-u^2/2}\, du \qquad (2)$$

the i's are certain functions of s and of y_1 and y_2 and the corresponding Z's, P's, and Q's, and δm, δs are increments to be added to provisional estimates m, s of μ and σ. The asymptotic variance-covariance matrix of the estimates is obtained as the inverse of the coefficient matrix calculated in the final cycle. The i's can be calculated directly from tables of Z and P, but the labour can be considerably shortened by expressing them either in terms of the incomplete normal moment function (Pearson, 1930, Table IX), the form in which they are given by Bliss and Stevens, or of the functions λ and ζ, defined and tabulated by the present author (**II**), when they are given by

$$\left.\begin{aligned} s^2 i_{\mu\mu} &= (1 - P_1 - Q_2) + P_1\lambda(-y_1) + Q_2\lambda(y_2) \\ s^2 i_{\mu\sigma} &= P_1\zeta(-y_1) - Q_2\zeta(y_2) \\ s^2 i_{\sigma\sigma} &= 2(1 - P_1 - Q_2) + P_1 y_1\zeta(-y_1) - Q_2 y_2\zeta(y_2). \end{aligned}\right\} \qquad (3)$$

2.4 *A modification of the Bliss-Stevens method*

The coefficients of (1) are expected values. They may be replaced, if convenient, by values calculated in terms of the observed sample, without affecting the final estimates. Garwood (1941) has shown that, in the analysis of quantal response data by the probit method, this replacement has the effect of reducing the number of cycles required for convergence. It seems likely that this result is generally true, and that this modification is therefore desirable, provided that it does not noticeably increase the labour of a single cycle. In the present case, in fact, using the tables of λ and ζ mentioned above, the labour, even of a single cycle, is slightly reduced, provided the method is also modified to give immediate estimates of α and β, rather than of μ and σ. In this notation the logarithm of the likelihood is

$$\log L = C + n_1 \log P_1 + n \log \beta - \sum_1^n (\alpha + \beta t_i)^2 + n_2 \log Q_2 .$$

When $T_1 = -\infty$ and $n_1 = 0$, the situation reduces to a particular case of the 'accidental death' model considered in the second paper of this series, and has already been discussed in that connection (**II**, §5.3). By a manipulation similar to that employed there, the iterative equa-

tions for the general case can be reduced to the form

$$\left.\begin{aligned} D_1 a_2' &= n_1\zeta(-y_1) - n_2\zeta(y_2) \\ D_2 b_2 &= n_1(T_1 - \bar{t})\zeta(-y_1) - n_2(T_2 - \bar{t})\zeta(y_2) + (2n/b_1) \end{aligned}\right\} \quad (4)$$

where

$$D_1 = n_1\lambda(-y_1) + n + n_2\lambda(y_2)$$

$$D_2 = \left\{(n/b_1^2) + n_1T_1^2\lambda(-y_1) + \sum_1^{n} t_i^2 + n_2T_2^2\lambda(y_2)\right\} - D_1\bar{t}^2$$

$$\bar{t} = \left\{n_1T_1\lambda(-y_1) + \sum_1^{n} t_i + n_2T_2\lambda(y_2)\right\} \Big/ D_1$$

$$y_j = a_1 + b_1T_j\,, \qquad (j = 1, 2)$$

and a_1 and b_1 are provisional estimates of α and β,

$$a_2 = a_2' - b_2\bar{t}$$

and b_2 are the revised estimates, and

$$\lambda(\eta) = \nu^2(\eta) - \eta\nu(\eta)$$

$$\zeta(\eta) = \nu(\eta) - \eta\nu(\eta),$$

where

$$\nu(\eta) = Z(\eta)/Q(\eta).$$

The final estimates a' and b are uncorrelated, and their asymptotic variances are estimated by the reciprocals of their coefficients D_1 and D_2 in (4), as calculated in the final cycle.

2.5 *Other methods*

Hald (1952) considers only truncation at one tail, and obtains a single maximum likelihood equation in α, which may be solved with the aid of a function $\psi'(x)(= \nu(-x))$, (Hald, 1952b, Table X). A second equation is then explicitly soluble for σ, and tables are given from which the variances and covariance of σ and $\mu = -\alpha\sigma$ can be obtained.

Cohen considers truncation at both ends of the range, and obtains maximum likelihood equations for α and σ, which are solved by an iterative process different from those used above. The variances and covariances of the estimates are then calculated separately.

All four methods here described are fundamentally maximum likelihood methods, and will thus give identical estimates for a given

sample. Which method is used in practice will depend upon ease of application, with or without tables, and rapidity of convergence. For single-ended truncation Hald's method is probably the most convenient, if the appropriate tables are available; if not, one of the three iterative procedures is probably preferable. Of these, that of Cohen is the most elegant, but the details of the solution are not explained in Cohen's paper, and the method is likely to be unfamiliar to the non-mathematician. Provided tables of λ and ζ are available, the method of §2.4 seems to be the most convenient, while if no tables are available other than those of Z and P, so that λ and ζ, or the i's of the Bliss-Stevens method, must be recalculated from these tables at each stage, there is probably little to choose between the three methods.

2.6 *Partial truncation*

This situation, in which the distribution function is truncated at different points for different individuals of the sample, is closely related to the 'accidental death' problem, and has already been discussed (II, §5.3) for single-tailed truncation. Simultaneous partial truncation at both ends of the scale seems rather unlikely, and I do not propose to discuss methods of analysis for this situation. Should such methods be required they can be obtained, by obvious generalisations, from those of §2.4 above and of II §5.2.

These methods should be applied with caution, as they depend on the assumption that the time of truncation is uncorrelated with the expected survival-time of any individual. For example, if the response-time under investigation is the duration of an artificially induced coma, different individuals will show different times of onset of coma. If observation is suspended at a given time, the period from onset of coma to time of truncation will vary from individual to individual. Such data may appear suitable for analysis by the 'partial truncation' method, but this will introduce a bias if, for example, those individuals with the earliest onset of coma (and so the latest 'time' of truncation) have an expected duration of coma greater or less than the mean duration for all individuals.

2.7 *Accidental deaths present*

If accidental deaths occur in the sample, all individuals surviving at the time of truncation may be considered as dying accidentally at that time, and the normal distribution estimated by the usual 'accidental death' method. The modification necessary in estimating the distribution of accidental deaths has already been discussed (II, §5.3).

2.8 *Compound of two normal distributions*

Estimation in the 'natural death' situation, when two stimuli, each of which, acting alone, would be expected to produce a normal response-time distribution, act simultaneously on a population, has been discussed in (**II**, §6). The joint p.d.f. is given by

$$h(t)\,dt = \{\beta_1 Z_1 Q_2 + \beta_2 Z_2 Q_1\}\,dt$$

where the suffices now refer to the two parent distributions, and Z_i and Q_i are functions of t defined as in (3), with y_i replaced by

$$\eta_i = \alpha_i + \beta_i t$$

Then if observation is suspended at time $t = T$, when n, n_T and n', n'_T, are the numbers which have responded and failed to respond in the control and treated series respectively, the log likelihood of the combined sample is given by

$$C + \sum_1^{n'} \log \{\beta_1 Z_1 Q_1 + \beta_2 Z_2 Q_2\} + n'_T \log Q_{2T} Q_{2T} + n \log \beta_2 - \frac{1}{2} \sum_1^{n} (\alpha_1 + \beta_2 t)^2 + n_T \log Q_{2T}\,.$$

The equations of estimation (**II**, §6.5) are modified by the inclusion of the two additional terms; the third equation there given becomes

$$\begin{aligned} \{\textstyle\sum WY + n'_T \nu_{1T} \nu_{2T}\}\ \delta\alpha_1 + \{\textstyle\sum XY + n'_T T \nu_{1T} \nu_{2T}\}\ \delta\beta_1 \\ + \{n + \textstyle\sum Y^2 + n_T \lambda_{2T} + n'_T \nu^2_{2T}\}\ \delta\alpha_2 \\ + \left\{\sum_1^n t + \textstyle\sum YZ + n_T T \lambda_{2T} + n'_T T \nu^2_{2T}\right\} \delta\beta_2 \qquad (5) \\ = -\left[n a_2 + b_2 \sum_1^n t\right] - \textstyle\sum Y - n_T \nu_{2T} - n'_T \nu_{2T}\,, \end{aligned}$$

where ν_{iT} is the value of ν corresponding to

$$\eta_{iT} = \alpha_i + \beta_i T,$$

and W, X, Y, and Z are the functions defined in (**II**). A comparison of (5) with its original form should indicate sufficiently clearly the modifications necessary to the other three equations. The form (5) appears, on paper, considerably more formidable than the corresponding form without truncation; the actual additional labour involved, however, is small compared with that required to calculate W, X, Y, Z and their sums of squares and products. It is assumed in the foregoing that the control and treated series are truncated at the same time;

the slight modifications necessary when the control series is truncated at T and the treated series at some time T' different from T are fairly obvious, and I shall not describe them in detail.

2.9 *Grouped data*

The analysis of grouped data has been discussed in (I, §3). No special methods for truncated data are necessary; if the grouping intervals, other than the truncated tails, are sufficiently small, a class mark may be attached to the observations falling in each group, and the data analysed as a straightforward truncated sample, ignoring grouping. Otherwise a special method for grouped data, such as that of (I, §3.7), must be used; such methods make no assumptions about the widths of the groups, so that the truncated tails are simply regarded as the extreme groups, and no modification is necessary.

2.10 *Calculation of provisional estimates*

For the simple normal distribution with single or double truncation, provisional estimates are mostly conveniently calculated by plotting empirical N.E.D.'s, obtained by transforming observed proportionate responses, against time for a few selected values of t in the observed range, and drawing a line by eye to fit the points. This line is then a provisional estimate of

$$\eta = \alpha + \beta t$$

and provisional estimates a_1 and b_1 of α and β can be read from the line. Cohen suggests calculating provisional estimates from the empirical N.E.D.'s for the two truncation points only: however, the introduction of two or three more points between them will not involve any great labour, and may considerably improve the provisional estimates.

If accidental deaths occur in the sample, this method will give adequate provisional estimates provided only a small proportion of the individuals die accidentally; otherwise the method used in the example of II, §5.5 can be used over that part of the range for which individual observations are available.

For the compound of two normal distributions, provisional estimates may be obtained as described in II §6.5 (i), and used directly in the maximum likelihood procedure of §2.10 above, omitting the estimation by moments suggested in II for non-truncated samples.

2.11 *Efficiency of estimation from truncated data*

A decision to suspend observation at any stage of an experiment can only be made on the basis of a knowledge of the amount of time

saved and the amount of information sacrificed by truncation. The loss of information on any parameter can be estimated by comparing the asymptotic variances of its maximum likelihood estimates calculated from a truncated and a completely observed sample; the ratio of these variances represents the efficiency of the truncation procedure relative to complete observation. The relative efficiencies, for μ and σ, of a singly truncated normal sample are given in Table I for degrees

TABLE I

PERCENTAGE EFFICIENCY OF TRUNCATED NORMAL SAMPLES FOR ESTIMATION OF μ AND σ.

	% individual responses observed						50% response; σ known	
	100	90	80	70	60	50	M.L. est.	Median
Eff (μ)	100	98	94	88	79	66	82	64
Eff (σ)	100	85	73	61	50	40	—	—
$\rho(\mu\sigma)$	0	0.053	0.125	0.214	0.320	0.441	—	—

of truncation up to 50%, together with the correlations of the estimates. (The variances and covariances of μ and σ for truncated samples, used in constructing this table, were theoretical values, calculated by inserting $\mu = 0$, $\sigma = 1$, $\eta_1 = -\infty$, and η_2 equal to the required percentage point, into the formulae (3), and inverting the i-matrix.) Halperin (1952) has given similar efficiencies for the estimation of μ only, with σ known; his figure for 50% truncation, and the relative efficiency of the median as an estimator for μ with σ known, are also included for comparison.

3. THE PROBLEM OF SURVIVAL

3.1 *Types of survival*

Even though observation is continued long after the last recorded reaction, there may remain some individuals that have not responded, and that can no longer be expected to do so. These individuals are said, by analogy with the particular situation in which the reaction is death, to *survive* the treatment *absolutely*. Two situations can be distinguished; that in which the individual remains, for some reason, totally unaffected by the stimulus, and that in which it is affected, but not to such an extent that it shows the reaction. Pursuing the analogy with time-mortality studies, I shall call these two types of survival *immunity* and *recovery* respectively.

When individuals surviving are 'immune', there is no reason to suppose that they are 'more resistant' than other individuals, in the sense that they would have survived longer than others had they not been immune: true immunity is to be regarded as a condition distributed quite independently of the expected survival-time. Thus the individuals surviving may be considered, at least for the purpose of response-time analysis, as a random sample of all individuals, and hence no bias is introduced by ignoring the surviving individuals, and analysing the observed response-times by standard methods. Immunity could occur, for example, in the following situations.

(a) When tumours are induced in mice by grafting small pieces of tumour material, some of the grafts may break down. The corresponding animals will not develop tumours, and in any study of survival-times or rate of growth of tumours, such animals must be considered 'immune'.

(b) In a study of incubation periods of infectious diseases in man, some members of a randomly selected sample may be immune as a result of acquiring or inheriting the appropriate antibodies.

(c) In a batch of eggs deposited by an insect, a proportion may be infertile. If incubation times are observed, these eggs can never show the expected reaction, and are thus 'immune', in the present sense.

When absolute survival is due to recovery, on the other hand, those individuals that recover will be those offering the greatest resistance to the operation of the stimulus, and these might be expected to survive longest if, say, the stimulus were applied at a greater intensity (for example, animals that recover from a moderate dose of a poison would probably have survival-times exceeding the average if a dose were given large enough to produce a 100% kill). The survivors are therefore by no means a random sample of all animals, but are drawn chiefly from the upper part of the response-time range, so that the distribution of observed response-times, while still normal for low values of t, will show a 'thinning-out' effect for large t. If cumulative kill, transformed to N.E.D.'s, is plotted against time, this thinning-out shows itself in a gradual falling-away from linearity with increasing time (c.f. Bliss & Stevens, 1937, Fig. 2).

In any particular case, both factors may operate, some individuals being truly immune, others recovering from the treatment.

There is a danger that the recovery-immunity problem may be over-simplified by discussing it in terms of death, rather than of a more general reaction. The difficulties are most easily seen by considering the effect of varying the intensity of the treatment. It may be argued that an individual that survives a small dose of a poison, but succumbs to a larger dose, is not truly immune. On this argument,

the proportion *immune* remains constant as the dose is increased, while the proportion *recovering* decreases. Fo. the straightforward poison-death system, where all individuals receive comparable treatment, this may well be true, at any rate for a restricted range of dose. However, for other reactions, the concept of 'recovery' becomes meaningless. An example is given by situation (c) above. A batch of eggs is observed; a certain proportion fails to develop. They are 'immune'. A comparable batch is kept, from the moment of deposition, at a much lower temperature than the first. The effect of this treatment will be to increase the incubation period, but it will also kill some of the eggs. Thus in this batch, not only is the average incubation time increased, but the proportion failing to develop (i.e. the proportion 'surviving') is greater than in the untreated batch. In some senses, the eggs sterilised by the treatment are different from those initially infertile. In a probit analysis, for example, the two groups would be distinguished, the first as the object of study, the second as a nuisance, to be allowed for by the application of a correction factor. However, in this case there is no such factor as that 'general level of resistance' which, in experimental animals, tends to correlate survival-time with recovery: there is no reason to suppose that the eggs sterilised by the treatment were those with the longest potential incubation period. In analysing the incubation-times, therefore, no distinction should be made between these eggs and those infertile when laid: both groups should be regarded as immune. A situation may thus arise in which, although the proportion surviving depends upon the 'dose' (or intensity of treatment) an assumption of immunity is justifiable for each individual dose. (If, however, the treatment is delayed until the eggs are partially developed, the probability that an egg will be sterilized may depend upon its degree of development, and so, indirectly, on its expected incubation period, so that the recovery model would be more appropriate.)

Again, when the strength of a bacterial suspension is assayed by the reaction of animals injected with small quantities of the suspension, the reaction will occur only if bacteria are present. Animals receiving no bacteria are then 'immune', in the present sense, and must be treated as such in the analysis of response-times, even though the 'proportion surviving' decreases with increasing dose or concentration of the suspension.

3.2 *A mathematical model for survival*

Any attempt to estimate the parameters of the response-time distribution presupposes a knowledge of the type of survival operating. A definite decision on this point can sometimes be made, either by

consideration of the mechanisms involved, as in some of the situations of §3.1, or on an observational basis, as when affected animals show some preliminary symptoms, but recover without showing the required reaction. In other cases, however, the decision may depend on the observed response-times themselves, either alone or in conjunction with other information of the type suggested above. In order to specify the hypotheses involved in a test of immunity against recovery, we must have some idea of the type of distribution of observed response-times to be expected when recovery operates, and of the manner in which this distribution departs from the normal form expected in the presence of immunity. A tentative model for recovery, leading to such a distribution, is put forward in this section.

Suppose that y is a measure of the tolerance of an individual, and that in the population y is jointly distributed with the potential response-time to any 'dose' (or intensity of treatment) in a bivariate normal form, with positive correlation ρ,

$$f(t, y) = \frac{\beta\zeta}{2\pi(1-\rho^2)} \exp -\frac{1}{2(1-\rho^2)} [(\alpha+\beta t)^2 - 2\rho(\alpha+\beta t)(\xi+\zeta y) + (\xi+\zeta y)^2]. \tag{6}$$

(Thus the more tolerant an animal, the greater its expected response-time.) Then if a 'dose' Y is applied, an individual survives absolutely provided

$$y > Y,$$

the probability of which is

$$\int_{-\infty}^{\infty} dt \int_{Y}^{\infty} dy\, f(t, y)$$

which, after some manipulation, reduces to

$$1 - \gamma = Q(\psi) \tag{7}$$

where ψ is a parameter depending on Y and the parameters of $f(t, y)$. The probability that an individual responds in the interval $t - (t + dt)$ is given by

$$\int_{-\infty}^{Y} dy\, f(t, y)\, dt$$

which reduces to

$$f_1(t)\, dt = \beta Z(\eta) Q(\kappa\eta - \psi\sqrt{1+\kappa^2})\, dt \tag{8}$$

where

$$\eta = \alpha + \beta t,$$

ψ is the parameter of (7), and

$$\kappa = \rho/\sqrt{1 - \rho^2}. \tag{9}$$

Thus if the response-time and the tolerance are uncorrelated, $\kappa = 0$, and the form (11) reduces, as expected for this 'immunity' situation, to

$$\gamma\beta Z(\eta)\, dt.$$

A perfect correlation between response-time and tolerance will result in simple truncation, an animal responding only if its expected response-time t is less than some fixed T. For this situation κ is infinite.

The function $f_1(t)$ is not, as it stands, a probability density function, inasmuch as its integral, over the whole range, is the probability $P(\psi)$ of obtaining a response. The required p.d.f. of *observed* response-times is therefore given by

$$f_1(t)/P(\psi). \tag{10}$$

This distribution has been discussed, in another context, by Birnbaum (1950). Its cumulant generating function is

$$\log \Psi(t) = \mu\theta + \tfrac{1}{2}\sigma^2\theta^2 + \log Q(K\theta - \psi)$$

where $\mu = -\alpha/\beta$, $\sigma = 1/\beta$, and $K = \kappa\sigma/\sqrt{1 + \kappa^2}$.
Thus the mean and second and third moments about the mean are

$$\left.\begin{aligned} \mu'_{1R} &= \mu - K\nu(-\psi) \\ \mu_{2R} &= \sigma^2 - K^2\lambda(-\psi) \\ \mu_{3R} &= -K^3\lambda'(-\psi), \end{aligned}\right\} \tag{11}$$

where ν, λ are, as elsewhere, the functions of **II**, §5.2.
The function $\lambda'(x)$ is positive for all x (Sampford, 1953), whence, since K is non-negative, μ_{3R} is negative for all ψ, except in the limiting case of immunity in which K, and hence μ_{3R}, are zero.

The most obvious test for recovery against immunity is thus the usual test, based on the third sample cumulant, for skewness against normality, considering only negative alternatives to the hypothesis, corresponding to immunity, $\kappa_3 = 0$ (Fisher, 1930). The insensitivity of tests based on third and higher order moments is notorious, so that a 'significant' result will seldom be obtained even with the one-tailed form of the test required here. (With a truncated sample and a survival-

rate of about 50%—conditions which produce the widest possible departure from normality—a sample of about 30 is needed to make Fisher's γ_1 equal to 1.64 of its standard deviation, and such favourable conditions will rarely occur in practice). However, a moderately large and negative third moment would strengthen a prior suspicion of recovery.

An alternative approach would be to test the conformity of the observed sample to normality with the χ^2 test. Here again, however, samples used in practice will normally be too small to detect differences of the order to be expected.

There are two dangers against which one must guard in applying the 'skewness' test. The first is that a non-normal distribution of response-times may arise not from recovery but from an inappropriate choice of time-metameter. The second is that the mechanism producing recovery might also delay death, so that the observed response-times in the upper part of the range, though fewer in number than would be expected on the hypothesis of normality, would also be greater than expected. This effect would tend to offset skewness caused by recovery, and might even lead to a positive third moment. This latter difficulty would be to some extent avoided by using the χ^2 test, rather than that based on the third moment.

It is clear from the above that we shall very often remain in doubt, even after all available information has been considered, whether in fact immunity or recovery is operating. In such a situation a decision for recovery seems the more satisfactory alternative, inasmuch as, while estimates based on the immunity hypothesis are inconsistent if recovery is operating, one of the methods (that of §3.4) of estimation on the recovery hypothesis gives estimates which, though biased, are at least consistent if this hypothesis is incorrect.

3.3 *Analysis with immunity present*

When the only factor tending to modify the normal distribution of response-times is the occurrence in the sample of immune individuals, the observed response-times can still, as noted above, be regarded as a random sample from the normal population, so that the mean and variance can be estimated from the observed values as from any other normal sample. The proportion immune, which is also necessary for a complete description of the population, can be estimated directly by the proportion surviving in the sample.

However, the methods for multi-stimulus distributions and truncated samples cannot be applied directly, as all these methods depend on the assumption that an individual killed or dying without showing

the reaction has a potential response-time conforming to the normal distribution, and greater than the time for which it has already survived; this is not true for immune individuals, and to assume it true would introduce a positive bias into the estimates of mean and variance. Special maximum likelihood methods can be constructed, but as, in general, they do not admit of any very marked simplification, I shall only outline their derivations.

(i) Truncation.

If observation is stopped at time T, while responses are still taking place, individual response-times will be available for all individuals responding before time T, say n in number, but there will remain n' individuals, any one of which either has a potential response-time greater than T, or is immune, and the difficulty arises from the fact that the two groups of survivors are indistinguishable. The simplest method of approach in this case is to regard the sample of individual response-times as a truncated sample with an unknown number of individuals in the truncated tail, and to analyse by one of the methods appropriate to this situation (Cohen, 1950; Hald, 1952). Once the parameters are estimated, the expected proportionate response at time T, P_T , may be calculated from them, and the proportion immune in the population, π_I , estimated by

$$P_I = 1 - \frac{n}{P_T(n + n')}$$

This method is not fully efficient, as it ignores the upper limit n' to the permissible number of individuals in the truncated tail; it is unlikely to be seriously inefficient when the proportion immune is large, but becomes increasingly so as this proportion decreases. Alternatively, we can construct a maximum likelihood procedure for the simultaneous estimation of the proportion immune and the parameters of the normal distribution. The equations are simply particular forms of those given below for the 'accidental death' model.

(ii) The accidental death model.

For a general discussion of this situation the reader is referred to II, §5. We assume here tnat a proportion $(1 - \gamma)$ of the members of the population are immune to the treatment, that the remaining proportion γ are capable of responding to the treatment, their response-times being normally distributed, and that individuals are likely to be accidentally killed, sacrificed, or otherwise removed from the sample before responding, such 'accidents' taking place in time according to a distribution function $F_2(t)$. Then:

Probability that an individual will respond in the interval $t - (t + dt)$

$$= \gamma\beta Z[1 - F_2(t)]\,dt$$

Probability that an individual will be killed, without showing the response, in the interval $t - (t - dt)$

$$= (1 - \gamma P)f_2(t)\,dt$$

Probability that an individual will survive to time T without responding

$$= (1 - \gamma P_T)[1 - F_2(t)]$$

where $Z = Z(t)$, $P = P(t)$ are values of the ordinate and cumulative distribution function of the normal distribution of response-times at time t, and $P_T = P(T)$.

Then if n individuals respond at times t_i, n' are accidentally killed at times t_i', and n'' are surviving, without having shown the response, when observation ceases at time T, the log likelihood takes the form

$$\log L = n \log \gamma + n \log \beta - \tfrac{1}{2} \sum^{n} (\alpha + \beta t)^2 + \sum{}' \log (1 - \gamma P) + A(t)$$

where $\sum'$ denotes summation over all individuals not showing the response, P having the value $P(t')$ for an individual accidentally killed and P_T for one surviving, and $A(t)$ is a function independent of α, β, and γ.

Assuming approximations a_1, b_1, c_1 to α, β and γ, the method of maximum likelihood leads to the equations for revised estimates (cf. II, §5.2):

$$a_2' \sum w_1 + \delta c_1 \sum{}' \phi_1 = - \sum{}' \zeta_{c1}$$

$$b_2\left\{\sum w_1(t - \bar{t})^2 + \frac{n}{b_1^2}\right\} + \delta c_1 \sum{}' t\phi_1 = \frac{2n}{b_1} - \sum{}' (t - \bar{t})\zeta_{c1}$$

$$a_2' \sum{}' \phi_1 + b_2 \sum{}' t\phi_1 + \delta c_1\left\{\frac{n}{c_1^2} + \sum{}' \psi_1^2\right\} = \frac{n}{c_1} + \sum{}' (y_1\phi_1 - \psi_1)$$

where b_2 and $a_2 = a_2' - b_2\bar{t}$ are revised estimates of β and α, and δc_1 is an increment to be added to c_1 to give a revised estimate c_2 of γ. In these formulae $\sum$ operates over all individuals, $\sum'$ is as defined previously, t is a general variate taking the values t_i, t_i', T, for the three groups of individuals, w_1 is a weight function given by

$$w_1 = 1 \qquad \text{for an individual responding,}$$

$$= \lambda_{c1} \qquad \text{for all other individuals,}$$

$$\bar{t} = \sum w_1 t / \sum w_1 ,$$

and the remaining functions are given by

$$y_1 = a_1 + b_1 t$$

$$Z_1 , P_1 = Z(y_1), P(y_1)$$

$$\nu_{c1} = c_1 Z_1/(1 - c_1 P_1)$$

$$\lambda_{c1} = \nu_{c1}^2 - y_1 \nu_{c1}$$

$$\zeta_{c1} = \nu_{c1} - y_1 \lambda_{c1}$$

$$\phi_1 = Z_1/(1 - c_1 P_1)^2$$

$$\psi_1 = P_1(1 - c_1 P_1)$$

The inverse of the coefficient matrix for the last cycle provides an estimate of the variance-covariance matrix of the final estimates.

(iii) The natural death model.

For a general discussion of this situation, see **II**, §6. We assume, as there, a normal distribution of treatment survival-times, applying here only to a proportion γ of the population, superimposed upon a normal distribution of natural survival-times, the cause of death being indeterminable in any individual. Then the combined distribution function of survival-times is given by

$$H(t) = 1 - (1 - \gamma P_1)Q_2$$

$$h(t) = \gamma \beta_1 Z_1 Q_2 + \beta_2 Z_2 (1 - \gamma P_1)$$

where the suffices 1 and 2 refer, respectively, to the distributions of treatment and natural survival times. The construction of the maximum likelihood equations proceeds on lines similar to those of **II** §6 (iii), and the equations will have a form resembling those derived there: they will differ in that the functions ν, λ, and ζ derived from distribution 1 will be replaced by ν_C , λ_C , and ζ_C as defined above, and in the occurrence of additional terms, and an additional equation, for the estimation of γ. For example

$$\frac{\partial}{\partial \alpha_1} \log L = -\sum \frac{\gamma \beta_1 \eta_1 Z_1 Q_2 + \gamma \beta_2 Z_1 Z_2}{\gamma \beta_1 Z_1 Q_2 + \beta_2 Z_2 (1 - \gamma P_1)}$$

$$= -\sum \left\{ \nu_{c1} - \frac{\beta_1 \lambda_{c1}}{\beta_1 \nu_{c1} + \beta_2 \nu_2} \right\}$$

and

$$\frac{\partial}{\partial\gamma}\log L = \sum \frac{\beta_1 Z_1 Q_2 - \beta_2 Z_2 P_1}{\gamma\beta_1 Z_1 Q_2 + \beta_2 Z_2(1-\gamma P_1)}$$

$$= \sum\left\{\frac{1}{\gamma} - \frac{\beta_2\nu_{2\xi}}{\beta_1\nu_{e1} + \beta_2\nu_2}\right\}$$

where

$$\xi = 1/\gamma(1-\gamma P_1).$$

3.4 *Analysis with recovery present—estimation by moments*

The model of §3.2 depends on the four parameters α and β, characterising the underlying normal response-time distribution, ψ, the N.E.D. of the probability of obtaining a response, and κ, measuring the correlation between potential response-time and capacity for survival. Although the value of κ is of some interest, we are usually more concerned with the other three parameters, particularly as estimates of κ are very imprecise.

An estimate U of ψ is obtained as the N.E.D. of the observed proportion responding, i.e. as the solution of

$$P(U) = n/N,$$

where N is the total sample size and n the number of individuals responding. This estimate is not of full efficiency, as ψ also appears in the p.d.f. of observed response times, but will usually be satisfactorily efficient. Its asymptotic variance is given by

$$V(U) = P(U)Q(U)/NZ^2(U)$$

$$= 1/N\nu(U)\nu(-U).$$

If now we calculate, from the observed response-times $t_1 \cdots t_n$, the quantities

$\bar{t}$; $s_R^2 = \sum (t-\bar{t})^2/(n-1)$; and $k_{3R} = n\sum (t-\bar{t})^3/(n-1)(n-2)$,

these three quantities are unbiased estimates of the moments given in (11). Setting them equal to their expected values, and substituting U for ψ, we have three equations which can be solved to give estimates of μ and σ^2,

$$m = \bar{t} + \tilde{K}\nu(-U)$$

and

$$s^2 = s_R^2 + \tilde{K}^2\lambda(-U)$$

where

$$\tilde{K} = [-k_{3R}/\lambda'(-U)]^{1/3}$$

(As this analysis assumes that survivors are recovering, not immune, k_{3R} will be negative and $\tilde{K}$ positive.) Should an estimate of κ be required, it can be calculated from $\tilde{K}$ and s.

Large-sample formulae for the variances and covariances of these estimates can be derived, but depend on moments up to the sixth order of the distribution of observed response-times, and are in consequence rather complicated. However, this distribution will not usually depart too seriously from normality, so we can derive approximate formulae by assuming that, for samples of size n,

$$V(s_R^2) = 2\mu_{2R}^2/n; \quad V(k_{3R}) = 6\mu_{2R}^3/n; \quad \text{cov}\,(\bar{t}, k_{3R}) = \text{cov}\,(s_R^2, k_{3R}) = 0.$$

Making these approximations,

$$\left.\begin{aligned} V(m) &= \frac{1}{n}\left\{s_R^2 + \frac{2\tilde{K}^2\nu^2}{3k_{3R}^2}\, s_R^6\right\} + \frac{\tilde{K}^2M^2}{9}\, V(U) \\ \text{cov}\,(m, s^2) &= \frac{1}{n}\left\{k_{3R} + \frac{4\tilde{K}^3\nu\lambda}{3k_{3R}^2}\, s_R^6\right\} + \frac{\tilde{K}^3ML}{9}\, V(U) \\ V(s^2) &= \frac{1}{n}\left\{2s_R^4 + \frac{8\tilde{K}^4\lambda^2}{3k_{3R}^2}\, s_R^6\right\} + \frac{\tilde{K}^4L^2}{9}\, V(U) \end{aligned}\right\} \qquad (12)$$

where

$$M = (2U\lambda^2 + \nu^2\zeta)/\lambda'$$

$$L = (\zeta^2 + 2\nu\zeta - 4\lambda^2)/\lambda',$$

$\zeta(x) = \nu(x) - x\lambda(x)$ is the function defined and tabulated in **II**, and ν, λ, ζ, and λ' are the values of these functions for $x = -U$.

3.5 *Estimation with recovery present—the artificial truncation method*

Bliss and Stevens (1937) suggested that, when individuals recover, the sample can be truncated at the point where the N.E.D. curve departs from linearity. All surviving animals are then supposed to have potential survival-times above this point, and the truncated sample is analysed by the method of §2.3 (or by one of the alternative methods of §§2.4-2.5). This approach ignores the possibility that some individuals have recovered whose potential survival-times were less than that chosen for the truncation point. Whether or not this possibility requires serious consideration depends on the character of the survival mechanism. If the model of §3.2 is valid the probability that some of the 'frustrated' values lie below the point of truncation may be quite large, and this may introduce an appreciable bias (see §§3.6-3.7). In the limiting case, however, when $\kappa = \infty$, or in more complicated situations, the probability of survival may be effectively

zero over much of the range of potential response-times, so that the risk of bias introduced by the choice of too high a point of truncation can be ignored.

3.6 *The point of truncation*

In applying the artificial truncation method, a point of truncation must be chosen below which the normal distribution may be assumed to operate satisfactorily. This point is usually selected by inspection of the observed N.E.D. line; it is of some interest to consider the determination of a suitable point from the model of §3.2.

Suppose the number of survivors to be n'. Each of these has an expected but unrealised response-time, and these 'frustrated' values have the p.d.f.

$$f_*(t)\,dt = \{Z(\eta)P(\kappa\eta - \psi\sqrt{1+\kappa^2})/Q(\psi)\}\,dt.$$

Then a point of truncation t_0 may be chosen to make the probability that one of the 'frustrated' values lies below t_0 equal to any small quantity ϵ, by writing

$$\left\{\int_{t_0}^{\infty} f_*(t)\,dt\right\}^{n'} = (1-\epsilon)$$

which is equivalent to

$$\frac{1}{2\pi\sqrt{1-\rho^2}}\int_{\eta_0}^{\infty} du \int_{\psi}^{\infty} dv \exp = \frac{1}{2(1-\rho^2)}[u^2 - 2\rho uv + v^2] = Q_\psi(1-\epsilon)^{1/n'}, \tag{13}$$

where $\rho = \kappa/\sqrt{1+\kappa^2}$ and $\eta_0 = \alpha + \beta t_0$.

Inserting into (13) provisional estimates of ψ and κ obtained as in §3.4, we obtain, with the aid of tables of the bivariate normal distribution (Pearson, 1931; Tables VIII-IX), an estimate y_0 of η_0 . The actual point of truncation can be determined either using the provisional estimates of α and β, or, preferably, by observing that the expected number of deaths occurring before time t_0 is

$$\begin{aligned} n_0 &= (n+n')P(\eta_0) - n'\{1-(1-\epsilon)^{1/n'}\} \\ &= nP(\eta_0) + n'\{(1-\epsilon)^{1/n'} - Q(\eta_0)\} \end{aligned} \tag{14}$$

and truncating at a point for which the number of retained observations is the nearest integer to the estimate of n_0 obtained by setting $\eta_0 = y_0$ in (14).

The requirement that none of the frustrated values should lie below the point of truncation may be felt to be unnecessarily severe,

and it will usually lead to the rejection of a rather large proportion of the available observations, so that in practice a subjectively determined truncation point, preserving considerably more than n_0 observations, might be preferred. The expected number of 'frustrated' values lying below any truncation point T can be calculated, by use of the bivariate normal tables, from the p.d.f. $f_*(t)$, provided provisional estimates of the parameters are available.

3.7 *Example*

The following constitute an artificial sample of t-values, constructed from the model of §3.2 with

$$\alpha = -3.0; \quad \beta = 2.0; \quad \kappa = 0.50; \quad \psi = 0:$$

Observed t: 0.56, 0.78, 0.95, 0.98, 1.00, 1.09, 1.16, 1.22, 1.23, 1.26, 1.28, 1.33, 1.44, 1.52, 1.58, 1.68, 1.78.

Number of survivors = 13.

(i) Estimation by moments

$$n = 17; \quad N = 30. \quad \text{Thus}$$

$$P(U) = 17/30 = 0.5667; \quad Q(U) = 0.4333.$$

$$U = 0.168.$$

$$\nu(-U) = 0.6941; \quad \nu(U) = 0.9078; \quad \text{whence}$$

$$V(U) = 0.052902$$

Also

$$\lambda(-U) = 0.5984; \quad \lambda'(-U) = 0.2371; \quad \zeta(-U) = 0.7946.$$

From the observed t-values

$$\bar{t} = 1.266; \quad s_R^2 = 0.102413; \quad k_{3R} = -0.006325.$$

Then

$$\tilde{K}^3 = 0.006325/0.2371 = 0.026675$$

$$\tilde{K} = 0.2988$$

whence

$$m = 1.226 + 0.2988 \times 0.6941 = 1.433$$

$$s^2 = 0.102413 + 0.1988^2 \times 0.5984 = 0.155839$$

$$s = 0.395$$

and

$$\tilde{K}/s = k/\sqrt{1 + k^2} = 0.756; \quad k = 1.15.$$

Also $M = 2.1218; \quad L = 1.2741,$ whence

$$V(m) \approx 0.0537; \quad \text{cov}(m, s^2) \approx 0.02337; \quad V(s^2) \approx 0.013380$$

$$\text{cov}(m, s) \approx 0.0296; \quad V(s) \approx 0.0214.$$

These quantities are calculated from the approximate formulae (12) of §3.4. Use of the more exact formulae, with higher order moments of the observed t calculated from the distribution, gave values:

$$V(m) = 0.0641; \quad \text{cov}(m, s) = 0.0330; \quad V(s) = 0.0213.$$

(ii) Estimation of α and β by the artificial truncation method.

Points of truncation were determined subjectively, by drawing the N.E.D. line of proportionate response and judging the point ($T = 1.45$) at which it appeared to depart from linearity, and by the method of the previous section, using the moment estimates of α, β, κ, and ψ. This method gave, with $\epsilon = 0.1$, truncation at $T = 0.99$, or after the fourth observation, and with $\epsilon = 0.25$, at $T = 1.10$, or after the sixth observation. As a matter of interest, artificial truncation estimates were made for these three points by the method of §2.4. The results are shown in Table II, together with the expected numbers of

TABLE II

EXAMPLE OF §3.7. COMPARISON OF ESTIMATES AND ESTIMATED VARIANCES

Method	Point of truncation T	m	s	$V(m)$	cov(m, s)	$V(s)$	Number frustrated $t < T$ Exp.	Obs.
Moments	—	1.433	0.395	0.0641	0.0330	0.0213	—	—
Artificial truncation	0.99	1.385	0.355	0.0463	0.0288	0.0228	0.1	1
Artificial truncation	1.10	1.413	0.371	0.0265	0.0128	0.0162	0.3	1
Artificial truncation	1.45	1.514	0.447	0.0119	0.0058	0.0100	2.7	5
Max. L.	—	1.434	0.387	0.0557	0.0350	0.0274	—	—
Complete Sample	—	1.375	0.408					
Theoretical values	—	1.500	0.500					

'frustrated' values lying below the point of truncation. The moment estimates are included for comparison, as are maximum likelihood estimates, which are too laborious to compute as a regular routine. The moment estimates are practically identical with the maximum likelihood estimates, and have variances of about the same magnitude, suggesting that the method of moments is of fairly high efficiency. The artificial truncation estimates increase with T, as would be expected, as more and more of the 'frustrated' t-values are incorrectly assumed to lie above T. The variances of these estimates, even for the extremely drastic truncation at 0.99, are smaller than those of the moment and maximum likelihood estimates, but, as has already been suggested, this increase in precision is, at least to some extent, misleading, resulting as it does partly from the unavoidable omission from the variance of a possibly important component.

The artificial sample used in these calculations was constructed from a complete normal sample of 30 values, removing as 'frustrated' those values which exceeded the corresponding members of another normal sample with parameters chosen to give the required values of κ and ψ. Estimates of μ and σ have been calculated from this complete sample, and these, together with the theoretical values and the actual numbers of frustrated values lying below the various truncation points, have been included in Table II for purposes of comparison.

3.8 *Analysis with both forms of survival present*

It may be possible, from observations made during the course of the experiment, to distinguish between individuals immune and those recovering. If this is so, the proportion immune in the population is estimated by that in the sample, and the distribution of response-times is estimated by the methods of §3.4-3.5, using only those animals responding and recovering.

If no distribution is possible between the two types of survival, the data may be truncated at a suitable point, and analysed by the method suggested for truncated data in §3.3(i).

3.9 *Estimation of functions of the parameters*

The proportion $p_{T'}$ killed at a given time T', or the time $t_{P'}$ required to reach a proportionate kill $P'(< P_{\psi})$ may be estimated from the fitted parameters.

(i) Immunity.

The estimated proportion killed at time T' is

$$p_{T'} = P_U P\left(\frac{T' - m}{s}\right) = P_U P(y_{T'})$$

with asymptotic variance estimated by

$$V(p_{T'}) = \frac{p_{T'}}{n}\left\{Q_U + \nu^2(-y_{T'})\left[1 + \frac{y_{T'}^2}{2}\right]\right\},$$

where m, s are the usual moment estimates of μ and σ.

The time required to reach a proportionate kill P' is given by

$$t_{P'} = m + sy'$$

where y' is defined by

$$P(y') = P'/P_U .$$

The asymptotic variance of $t_{P'}$ is given by

$$V(t_{P'}) = \frac{s^2}{n}\left\{1 + \frac{y'^2}{2} + \frac{Q_U}{\nu^2(-y')}\right\}.$$

(ii) Recovery.

If the artificial truncation method has been used, approximate estimates of $p_{T'}$ and $t_{P'}$ can be made for the lower part of the range, over which the N.E.D. lines is assumed linear. The methods used for estimating $p_{T'}$ and $t_{P'}$ and their variances have already been discussed in I, §4.

If the method of moments has been used, leading to estimates of α, β, κ and ψ, the proportion killed at time T, or the time required to produce a proportionate kill P, can be estimated with the aid of tables of the bivariate normal distribution. Formulae for the asymptotic variances of these estimates can be derived, but are tedious to compute.

3.10 *Conclusion*

Which of the methods here described for the analysis of data with recovery present is used in any one case seems to depend on the amount of reliance which can be placed on the model of §3.2. If this model is accepted, the moment method is preferable, as it avoids the possible bias introduced by an inappropriate choice of truncation point, and will give consistent, though not unbiased, results if applied to a sample in which the survivors are, in fact, immune. The artificial truncation method, on the other hand, has the advantage of being much less dependent on any assumed form of recovery mechanism, and may therefore be preferred when the present model is considered inadequate.

4. APPLICATIONS TO BIOLOGICAL ASSAY

4.1 *Time as a variate in bio-assay*

When different doses of a poison are given to a series of groups of animals, the mean survival-time will in general depend on the dose: this dependence can often be expressed, by the choice of convenient metameters for time and dose, as a linear regression relationship (Box and Cullumbine, 1947). This suggests the use of survival-time as a quantitative measure of potency for use in biological assay, an application which has been discussed by Perry (1950) and Finney (1952; Ch. 22).

In comparing two dilutions of an experimental material, equivalent concentrations should produce the same mean survival-time. If, therefore, the dose metameter, x, is the logarithm of the dose, the regression lines for the two preparations will be parallel, and can be estimated as

$$t_1 = a_1 + bx$$

$$t_2 = a_2 + bx,$$

and the relative dosage value can be estimated as

$$m_{12} = \frac{a_2 - a_1}{b} = \frac{\bar{t}_2 - \bar{t}_1}{b} - (\bar{x}_2 - \bar{x}_1).$$

Then if t is normally distributed for each value of x, a_1 , a_2 , and b are also normally distributed, and fiducial limits for m_{12} may be obtained by the usual formula for limits to the ratio of two normal variates.

It is assumed that the assay is carried out at doses sufficiently large to ensure the death of all the animals. Then if the distribution of response-times assumed for a dose x of preparation j is

$$\frac{\beta \exp\left(-\tfrac{1}{2}\{\beta t + \alpha_j + \gamma x\}^2\right)}{\sqrt{2\pi}}\, dt \qquad j = 1, 2,$$

the asymptotic variance of the relative dosage value is given by

$$V(m_{12}) = \frac{1}{\gamma^2}\left\{\frac{1}{\sum_i n_{1i}} + \frac{1}{\sum_i n_{2i}} + \frac{(\bar{x}_2 - \bar{x}_1 + m_{12})^2}{\sum_j \sum_i n_{ji}(x_{ji} - \bar{x}_j)^2}\right\} \qquad (15)$$

where n_{ji} is the number of animals treated with preparation j at dosage x_{ji} .

As an alternative to the regression of time on dose, the experiment could be stopped at some time T, and estimates calculated from the proportions killed by each dose up to that time. These proportions

will be given by the probit lines

$$\eta_j = (\alpha_j + \beta T) + \gamma x \qquad j = 1, 2,$$

and the asymptotic variance of a relative dosage value estimated from these lines (if the heterogeneity χ^2 is not large)

$$V(m'_{12}) = \frac{1}{\gamma^2}\left\{\frac{1}{\sum_i n_{1i}w_{1i}} + \frac{1}{\sum_i n_{2i}w_{2i}} + \frac{(\bar{x}'_2 - \bar{x}'_1 + m'_{12})^2}{\sum_j \sum_i n_{ji}w_{ji}(x_{ji} - \bar{x}'_j)^2}\right\} \quad (16)$$

where the w's are the weights used in probit analysis. The weighted means $\bar{x}'$ of (16) will usually be rather different from the unweighted means $\bar{x}$, of (15), but the principal difference between the two forms is the appearance of the weights in the denominators in (16). As these weights are never greater than 0.64, and may be much less, the variance defined by (16) will clearly be considerably greater than that defined by (15). Some compensation may be provided by a considerable saving in time. For a given experiment, $V(m'_{12})$ is a minimum if T is chosen so that the proportions killed at the different doses are distributed more or less evenly on either side of 50%. With a mean survival-time of the order of several days, or longer, it might be economically desirable to perform a daily analysis on proportions killed, stopping the experiment as soon as m_{12} is estimated with the required degree of precision. If this has not been attained by the time $V(m'_{12})$ reaches its minimum, (roughly speaking, when one half of all animals are dead), the experiment can be carried through to completion, and m_{12} estimated from the regression of t on x. The methods of §2 could be extended to estimate α, β, and γ from response-time data truncated at time T, but the calculations would be laborious, and probably uneconomic unless the duration of the experiment were measured in weeks, rather than days.

4.2 *Comparison with the quantal response method*

In the argument of the previous section, the 'proportions surviving', on which the assay is based, are artificial, being produced by truncation, with the result that the N.E.D. regression line of proportionate survival on dose can be expressed in terms of the parameters of the distribution of survival-times. The alternative, and more usual, form of quantal response assay is carried out at doses low enough to give a considerable 'absolute' survival-rate. In this case, the N.E.D. line depends not on the time-mortality distribution, but on the relation between dose and absolute survival; the slope, γ', of the line may be greater or less than the slope γ used in the previous section. The formula for the variance

of the relative dosage value may be obtained from (16), by replacing γ by γ'. It is clear that γ' would need to be very much greater than γ, to make an assay of this type as efficient, statistically, as a complete time-response assay using the same number of animals, and having the x-values similarly distributed about their mean. The latter type of assay requires larger quantities of the preparations being assayed (not, in general, a serious difficulty) and more careful supervision; on the other hand, the use of larger doses may result in a shortening of the total duration of the experiment.

Perry (1950) has recommended the use of response-times in bioassay, on grounds similar to those discussed above.

4.3 *Additional information from response-times in a probit analysis*

There is, of course, no reason why the individual response-times, as well as the proportion surviving, should not be observed in a quantal response assay. An analysis of these survival-times, however, would then involve the estimation of the regression of time on dose, with the occurrence of survival.

If the response-times can be analysed on the assumption of immunity for each dose separately, no difficulty arises. (This situation is not, unfortunately, very common: however, some examples are suggested in §3.1). The mathematical model is obtained from formulae (7) and (8), setting $\kappa = 0$ and replacing ψ by $\psi + \varphi x$, and α by $\alpha + \gamma x$. The likelihood factorises into two independent parts, and (ψ, φ) and (α, β, γ) can be estimated by, respectively, a probit analysis and a linear regression of time on dose. A maximum likelihood estimate of m_{12}, the relative dosage value of two preparations, may be obtained by replacing x by $(x - m_{12})$ in the likelihood of the second sample, and maximising for ψ, φ, α, β, γ and m_{12}. This method, however, is rather clumsy. Alternatively, separate estimates of m_{12} may be made from the probit analysis and the time regression, and a weighted mean of the two values calculated. The slight loss of efficiency involved in this process is a small price to pay for so considerable a simplification of the calculations involved. Fiducial limits to m_{12} can be calculated, if required, (§4.4).

In most experiments of this type, however, the recovery, rather than the immunity, hypothesis, is appropriate. A complete mathematical model could be constructed, by extension of equations (7) and (8), and a maximum likelihood method constructed. Alternatively, the method of artificial truncation could be applied to the response-times at each dose separately, and an extension of one of the methods of §§2.3-2.5 used to give an estimate of m_{12} based only on the time

regression; the final estimate would then be taken as a weighted mean of this value and that obtained from the probit analysis. However, for both these methods the computations required would be too arduous for anything approaching routine use.

In practice a less refined method may be adequate. The effect of recovery on an initially straight regression line with uniform variance can be seen by considering the moment formulae (11) of §3.2. For high doses, with kills nearing 100%, the mean and variance of the response-times will tend to those for the original line. As the proportionate recovery increases, however, the mean will fall below the original line, and the variance about the line will decrease. The actual regression line will be curved, but neither this curvature nor the change in the residual variance is likely to be serious except for doses having large recovery rates. With a satisfactorily designed assay, such that corresponding doses of the two preparations produce roughly the same response, reasonably reliable estimates of both the relative potency and its variance should result from a regression of 'time' on 'dose', using only the response-times for those doses producing more than a suitable proportionate kill, which should, ideally, be chosen beforehand in the light of all available information on the likely value of κ—the smaller this value, the fewer the number of doses that need be rejected.

4.4 *Fiducial limits.*

When parallel probit lines are fitted to two series of data, the relative dosage value of the two treatments is estimated by

$$m_{12} = \frac{(\bar{y}_1 - \bar{y}_2) + b(\bar{x}_2 - \bar{x}_1)}{b}$$

where the numerator and denominator are (asymptotically) normal variates, whose variances and covariances have been estimated. Fiducial limits to m_{12} may be estimated by the usual method for the ratio of two normal variates (Fieller, 1940).

If parallel regression lines of time on dose

$$t_i = a_i + b'x, \qquad i = 1, 2 \tag{17}$$

have also been calculated, an estimate m_{12}' of the relative dosage value is given by

$$m_{12}' = (a_2 - a_1)/b',$$

and this will be independent of the probit estimate whether the lines (17) have been calculated by an unweighted linear regression or by

the artificial truncation model. In the former case m'_{12} is the ratio of two normal variates; in the latter case it is the ratio of two quantities which may be assumed asymptotically normal, by virtue of the limiting properties of the maximum likelihood procedure. Fiducial limits may thus also be constructed for m'_{12} .

Asymptotic variances can be calculated for m_{12} and m'_{12} , and used to construct a weighted mean, and provided b and b' are considerably greater than their standard errors, a variance can be calculated for this mean. Preferable to this procedure, however, is the direct calculation of fiducial limits to the relative dosage value, utilizing the information from both sets of observations (Finney, 1952; Ch. 2).

5. CONCLUDING REMARKS

In this series of three papers I have attempted to give a reasonably complete account of the more important complications occurring in the analysis of response-time data, with an account of some methods of analysis suitable for the treatment of data in which these complications are manifested. In an attempt to fill in some of the more obvious gaps in the subject I have described several new computing routines, and advanced a considerable number of tentative suggestions. Some of these routines are easily carried out; others are far more complicated, and unlikely to be of more than occasional practical use. In particular it may be felt that in the present paper I have built rather too freely on hypotheses which may represent a mathematical oversimplification of actual situations.

This may well be so, and, indeed, although the models used here may prove to be of some practical value, it would be unreasonable to expect them to fit all observed situations. However, only an empirical knowledge of such situations, based on far more extensive records than are at present available, can indicate whether these models are of general applicability and, if not, how they should be modified.

APPENDIX

A reference guide to the methods of analysis for various situations

This series of three papers has described methods for the estimation of the parameters of the normal response-time distribution in the presence of various complicating effects, the methods having been grouped under the three headings 'General', 'Multi-stimulus distributions' and 'Truncation and Survival'. This classification, however, is oversimplified; any given sample may show complications of more than one type, and methods developed for one situation are sometimes appli-

cable to other. The following summary of situations discussed, with reference to the relevant sections, is intended not as an exhaustive index, but as a guide to the methods appropriate to a particular situation.

A. Sample complete.

1.	Ungrouped data	I, §2.2
2. (a)	Regular grouping	I, §3.2
(b)	Transformed regular grouping	I, §§3.4-3.5; III §2
(c)	Irregular grouping	I, §§3.6-3.7

B. Sample incomplete: observations deliberately omitted.

1. (a)	Truncation	III, §§2.3-2.5
(b)	Truncation in grouped samples	III, §2.9; I, §§3.3, 3.7
2.	Partial truncation	II, §§5.2, 5.3(i); III, §2.6

C. Sample incomplete as result of removal of subjects by uncontrollable but identifiable causes ('Accidental death' model).

1.	General	II, §5
2.	Subjects accidentally killed	II, §5.3(iv)
3.	Subjects removed according to second normal distribution (natural deaths, or a second applied treatment)	II, §5.3(v)
4.	More than two distinguishable causes.	II, §5.3(vi)

D. 'Natural death' situation (same response produced by two stimuli, each producing normal response-time distribution—impossible to determine stimulus responsible in any particular case).

(a)	General	II, §§6.5, 6.7
(b)	Grouped data (See also E1(d) below)	II, §6.8

E. Sample incomplete as result of absolute survival.

1. (a)	Survivors immune	III, §3.3
(b)	Survivors immune—sample truncated	III, §3.3(i)
(c)	Survivors immune—accidental deaths occurring	III, §3.3(ii)
(d)	Survivors immune—natural deaths occurring	III, §3.3(iii)
2.	Cause of survival doubtful	III, §3.4
3.	Survivors recovering	III, §§3.4-3.6
4.	Both forms of survival present	III, §3.8

REFERENCES

Birnbaum, Z. W. Effect of linear truncation on a multinormal population. *Ann. Math. Statist.*, *21*, 252, 1950.

Bliss, C. I., and Stevens, W. L. The calculation of the time-mortality curve (Appendix—The truncated normal distribution). *Ann. Appl. Biol.*, *24*, 815, 1937.

Cohen, A. C., Jr. Estimating the mean and variance of a normal distribution from singly truncated and doubly truncated samples. *Ann. Math. Statist.*, *21*, 557, 1950.

Cohen, A. C., Jr., and Woodward, J. Tables of Pearson-Lee-Fisher functions of singly truncated normal distributions. *Biometrics*, *9*, 489, 1953.

Finney, D. J. *Statistical Method in Biological Assay.* Charles Griffin & Co. Ltd., 1952.

Garwood, F. The application of maximum likelihood to dosage-mortality curves. *Biometrika*, *32*, 46, 1941.

Gupta, A. K. Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika*, *39*, 260, 1952.

Hald, A. *Statistical Theory with Engineering Applications.* John Wiley and Sons, Inc., 1952a.

Hald, A. *Statistical Tables and Formulas.* John Wiley and Sons, Inc., 1952b.

Halperin, M. Maximum likelihood estimation in truncated samples. *Ann. Math. Statist.*, *23*, 226, 1952.

Pearson, K. *Tables for Statisticians and Biometricians, Vol. I.* 3rd edition. Cambridge University Press for the Biometrika Office, London, 1930.

Pearson, K. *Tables for Statisticians and Biometricians, Vol. II.* Cambridge University Press for the Biometrika Office, London, 1934.

Perry, W. L. M. Reports on biological standards VI. The design of toxicity tests. *Spec. Rep. Ser. Med. Res. Coun., Lond.*, no. 270, 1950.

Sampford, M. R. The estimation of response-time distributions. I. Fundamental concepts and general methods. *Biometrics*, *8*, 13, 1952a.

Sampford, M. R. The estimation of response-time distributions. II. Multi-stimulus distributions. *Biometrics*, *8*, 307, 1952b.

Sampford, M. R. Some inequalities on Mill's ratio and related functions. *Ann. Math. Statist.*, *24*, 130, 1953.

Walsh, J. E. Some estimates and tests based on the r smallest values in a sample. *Ann. Math. Statist.*, *21*, 386, 1950.

QUERIES

GEORGE W. SNEDECOR, *Editor*

112 **QUERY:** An experiment on chinch bug control was carried out in an ordinary randomised block design. There were eight treatments and four replicates. For each experiment a metal cylinder was stuck in the sand in each plot and kept partly filled with water and the number of chinch bugs seen in five minutes counted. The results are given in Table 1.

TABLE 1.

Treatment / Replicate	1	2	3	4	5	6	7	8
I	0	0	0	1	2	2	0	6
II	0	2	0	0	0	1	5	3
III	0	0	7	1	3	4	4	6
IV	0	0	3	1	4	14	10	9

An ordinary analysis of variance gives the following:

	SS	DF	MS	
Treatments	151.0	7	21.6	$F = 3.1^{\circ}$
Replicates	76.5	3		
Remainder	144.6	21	6.9	
Total	382.1	31		

The treatments are significant at the 5% level. However, if an attempt is now made to separate out the treatments by means of significant differences, none of the treatments appears to be significantly different from the others. Why this apparent inconsistency?

ANSWER: In your example, the F test is just barely significant. The requirements for significant differences by various methods need not be quite the same, and in marginal cases we may expect two methods to disagree, sometimes in one direction and sometimes in another. Your data is such an example.

Instead of discussing the data further in its present form, we feel that

its appearance suggests that a transformation should be considered. The difference between replicate I and replicate IV is surely larger for treatments 6, 7 and 8 than for treatments 1, 2 and 4. This variation in replicate difference seems associated with the variation in replicate mean. The elimination of this effect, if possible, by a transformation would serve to make the experimental results more clear cut.

One result which may flow from the utilisation of a transformation is that the data may now exhibit more nearly the property of 'additivity', that is, the effect of rows and columns are purely additive, whereas earlier it did not do so. A check on this may be made by isolating from the 'Remainder' sum of squares in the ordinary cross classification analysis of variance one degree of freedom labelled 'non-additivity'. The isolation and testing of this quantity for significance or otherwise goes as follows (See also J. W. Tukey, Biometrics 1949, 5, p. 232).

Round the borders of the table we calculate the deviations of the row and column means from the grand mean. Then we take the quantity

$$\sum \text{Cell value} \times \text{Row mean deviation} \times \text{Column mean deviation}$$

where the summation extends over all the cells in the table. The final result, θ say, has the sign of the apparent non-additivity effect. To convert this into a sum of squares we take

$$\theta^2/[\sum(\text{Row mean deviations})^2]\,[\sum(\text{Column mean deviations})^2]$$

and this will have one degree of freedom. Its ratio to the error mean square will be an F or 1 and 20 degrees of freedom, and the square-root of this ratio, to which we give the sign of θ, will have a t distribution with 20 degrees of freedom.

Carrying this out for our particular case we get

	SS	DF	MS	
Treatments	151.0	7		
Replicates	76.5	3		
Non-additivity	59.9	1	59.9**	$t = +3.77$
Balance	84.7	20	4.2	
Total	382.1	31		

This value of t is very significant, indicating a definite deviation from additivity in the effects of treatments and replicates.

your chinch bug example), the semicircles for

$$\log (x + c), (x + c)^{-1/2}$$

and other transformations with $p \leq 0$ are misleading. As c tends to zero, all these transformations (for $p \leq 0$) tend to (a translated multiple of) the very simple transformation:

$$\begin{cases} \text{score} = 0 & \text{if } x < 0, \\ \text{score} = 1 & \text{if } x \geq 0. \end{cases}$$

This has been indicated by the broken lines in Figure 1.

In Figure 2 the value of t for the different forms of transformation are given against the appropriate point in the transformation picture.

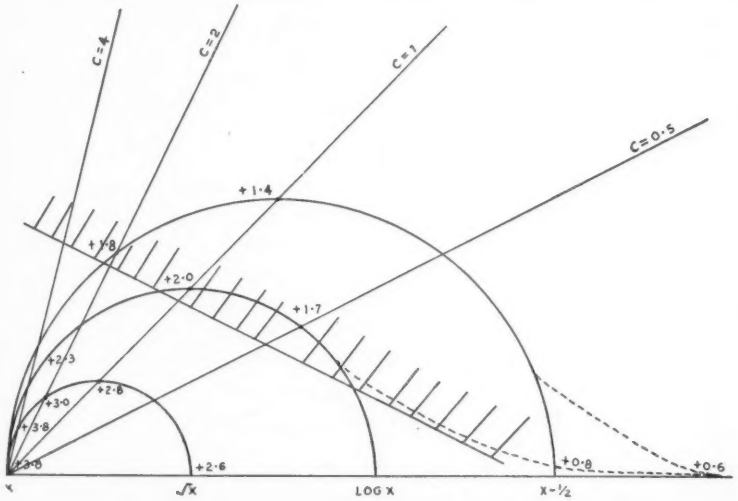


FIGURE 2.

VALUES OF t FOR NON-ADDITIVITY

All the values of t are positive. Since t for this particular problem has 20 degrees of freedom the 95% significance point is 2.09. If there is any one of these transformations for which the chinch bug experiment behaves additively, then in 95% of such experiments $-2.09 \leq t \leq 2.09$ for this transformation. Hence the region of the transformation quadrant in which $-2.09 \leq t \leq 2.09$ is a 95% confidence region for the additive transformation (if there be any such). The values in Figure 2 are numerous and well enough placed to allow us to draw, roughly, the curve

along which $t = 2.09$ and then to shade the area where $-2.09 \leq t \leq 2.09$. This we have done.

With more extensive data, possibly from a larger experiment, but more usefully from a number of experiments, we may find out quite accurately what transformation is appropriate. For the chinch bug data we have ruled out x itself, all $(x + c)^{1/2}$, and large values of c for $\log (x + c)$ and $(x + c)^{-1/2}$. We still have a wide choice left.

In Figure 3 the ratio of

$$F = \frac{\text{Mean square for treatments}}{\text{Mean square for interaction}}$$ is plotted.

This ratio varies slowly and not very far for this set of data. In other cases, wider and more rapid variation of this ratio might indicate which of the transformations not excluded by manifest lack of additivity seemed

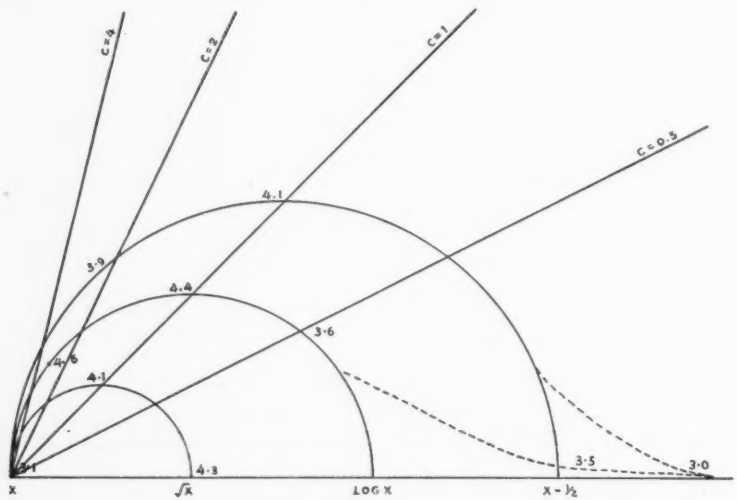


FIGURE 3.
COLUMNS/INTERACTIONS MEAN SQUARES

desirable. (With only a small body of data available, the problem of the effect of selection on significance would of course arise.) In the present example, there is no strong indication of a preference.

Combining the confidence region of Figure 2, the vague indications of Figure 3, and a natural, but doubtfully supported, desire not to go "too far" from the original variable, we are led to $\log (x + 1/2)$ or $\log (x + 1)$ as the desired transformation.

The analysis of variance for log $(x + 1/2)$ is shown in Table 2.

TABLE 2

Source	SS	DF	MS
Treatments	3.512	7	0.502**
Replicates	1.093	3	0.364
Non-additivity } error	.370 } 3.083	1 } 21	0.147
Balance } error	2.713 }	20 }	
Total	7.689	31	

The treatments are thus significantly different at the 1% level of significance. The next stage is to sort out the various means and see whether we can group the treatments together in some way. Ranked in order the means, of the transformed variates, are

$$-.301, -.124, .057, .204, .323, .528, .597, .787.$$

Following the method given by J. W. Tukey (Transactions of the New York Academy of Sciences, 1953, Vol. 16, p. 88) the first quantity to find is the WSD (wholly significant difference). To do this we find the total range in the table of transformed observations by taking differences in one direction and the ranges of the differences in the other direction and finally summing the ranges. This comes to 7.84 for this case and from Table 3 of the paper quoted above the appropriate factor is

$$\frac{\text{WSD for Totals}}{\text{Total range of differences}} = 0.42.$$

Hence WSD for means comes to $7.84 \times 0.42/4$ or 0.82. Thus, we can accept all the observed differences (as mean differences of log $(x + 1/2)$ scores) among treatments as correct ± 0.82 with only a 5% risk of being wrong about 1 or more differences.

In particular, the difference between treatment 8 and treatment 1 is found to be

$$0.79 - (-0.30) \pm 0.82$$

or $$1.09 \pm 0.82 = 0.27 \text{ to } 1.91$$

This is always positive so there is a definite difference between the two. A similar situation holds with the following pairs of treatments; 1 and 6, 1 and 7, 2 and 8.

In view of the fact that the analysis based on Figure 2 gave no choice between log $(x + 1/2)$ and log $(x + 1)$ for the chinch bug data, it is of interest to compare the results of the two transformations. This is done in Table 3.

TABLE 3

Treatment	Treatments means for:	
	log $(x + 1/2)$	log $(x + 1)$
1	−.30	.00
2	−.12	.12
3	.20	.38
4	.06	.22
5	.32	.44
6	.60	.66
7	.53	.63
8	.79	.82
WSD	±.82	±.63
Differences of established sign at 5%	$8 \geq 1, 2$ $7 \geq 1$ $6 \geq 1$	$8 \geq 1, 2$ $7 \geq 1$ $6 \geq 1$

As soon as we allow for the differing sizes of the WSD's, we see that we have learned almost the same things by using either transformation in the analysis.

PETER G. MOORE AND JOHN W. TUKEY

ABSTRACTS

Australasian Region

298 P. A. MORAN. **The Dilution Assay of Viruses.**

Dilution assays of viruses with eggs often do not fit the exponential curve given by the zero term of the Poisson distribution. When this happens the dilution assay is invalid as an assay. The cause is the variation in infectibility between eggs and it is important to be able to recognise when this happens. A very rapid test is given for the fitting of a dilution series to an exponential curve. This is believed to be more powerful than a χ^2 test and cuts the time of calculation by a factor of about 30.

299 H. O. LANCASTER. **Significance Levels in Discrete Distributions.**

In a continuous distribution a variable, t say, has a certain distribution. In a proportion of cases, α, t will be greater than $t\alpha$. α is said to be the significance level. If $t > t\alpha$, t is said to be significant at the α-level or the $100\alpha\%$ level. Under the null hypothesis the effective significance level will also be α. In discrete distributions the inequality is still used but the nominal and effective significance levels no longer agree. It is easy to find cases where the effective significance levels may be 2% and the nominal 5%. Such differences lead to great loss of power of the test. This whole problem is closely linked to the combination of probabilities from different experiments. A possible solution is to use the median probability as the test of significance and for the combination of experiments. This median probability usually closely corresponds with the probability of the uncorrected χ^2. This approach can be harmonised with the Stevens-Tocher auxiliary random sampling methods.

300 F. E. BINET AND H. T. CLIFFORD. **A Quantitative Study of an Assumed Hybrid Swarm between Two Eucalyptus Species.**

This paper describes the statistical analysis of three measured characters of fruit and peduncles, sampled from each of fifty trees of a suggested hybrid swarm, and from the same number of trees in each of the parental species (*E. elaeophora* and *E. goniocalyx*).

Appropriate multivariate tests of significance (including those applied in generalised discriminant analysis) support a model in which the

transformed characters form three trivariate normal distributions, with identical dispersion matrices, their mean vectors defining collinear centroids, the centroid of the hybrid population situated equidistant from those of the parents.

Estimated rates of misclassification are given for the discriminant function based on this model. A mating system is suggested, by which a hybrid swarm conforming to this model could have arisen.

301 J. D. BIGGERS AND P. J. CLARINGBOLD. **The Angular Transformation in Quantal Analysis.**

I. The angular transformation has been used in two ways for the analysis of quantal data:

(1) In the equalization of variance. Here the percentage data are transformed to angles and a direct analysis is carried out on these values.

(2) In an exact maximum likelihood solution. This follows the usual methods of probit analysis in which expected angles and working angles are used to give an iterative solution.

The two methods will be discussed in relation to the analysis of factorial experiments.

II. In factorial experiments in which the response is in quantal form many treatment combinations give groups of 0 or 100% response. By the use of treatment combinations which have an expected response between 5 and 95% these regions of extreme response may be avoided and the angular transformation used in the analysis. These treatment combinations are said to fall into the *region of useful observation* and give rise to skew experimental designs. Examples of these designs will be given.

THE BIOMETRIC SOCIETY

Interests of Members of the Society. The first Directory of the Society in 1949 listed the interests of our members. These were analyzed by Professor John W. Tukey of Princeton and reported in BIOMETRICS the following year (Volume 6, pages 198–199). Our second Directory, published four years later, carried a similar listing. To facilitate a more detailed analysis, these interests and other pertinent information concerning each member have been coded and punched on IBM cards through the kindness of Dr. E. C. Hammond and the American Cancer Society. The project has been directed by Dr. Colin White of Yale University, Assistant Secretary-Treasurer of The Biometric Society, to whom we are most grateful for the following report.

When the 1953 Directory was prepared, The Biometric Society consisted of 1,144 members. Their interests, as reported in the Directory or as inferred from their positions, have been classified, and the numbers in each classification counted. If a member reported two or three interests, each one was counted as one-half or one-third, respectively; and for convenience in the use of punched card equipment, any interests over three were ignored. The entries for two members (in the Eastern North American Region) defied classification.

A glance at the Directory will show some of the shortcomings of these data, which were collected in answer to an "open-ended" question. For instance, the Society includes many of the great teachers of statistical methods, and a large proportion of the members must do some teaching, but very few, in fact only 4.2 out of 1,144, have claimed teaching as an interest. Or again, an author of one of the major texts on sampling does not have sampling included in his "interests".

For the Society as a whole, the main interests have been expressed as a percentage of the total and within each section subdivided in terms of the number of member interests. Each member interest, of course, usually represents the partial interests of a larger number of members.

32.4% (a) Statistical sciences (370.3 as follows: in biology 106.1, in experimentation 45.5, in industry 39.0, bioassay 22.0, in agricultural research 21.0, in medical research 18.5, in sampling 15.3, other methodology 14.1, theoretical statistics 8.7, quality control 7.2, psychometrics 6.7, other and unspecified 66.2)

16.8% (b) General biology (192.3 as follows: genetics 99.2, ecology 31.0, general physiology 17.0, growth 13.8, zoology 11.5, theoretical biology 7.2, botany 4.0, unspecified 8.6)

12.9% (b) Applied biology and agriculture (147.7 as follows: nutrition 25,0, agricultural crops 21.7, agricultural land 19.2, forestry 18.0, animal mana ement 17.7, entomology 14.0, wildlife and fisheries 12.2, plant pathology 7.5, other and unspecified 12.4)

10.8% (c) Sciences usually related to medicine (123.0 as follows: pharmacology 58.0, microbiology 17.2, biochemistry 14.8, endocrinology 12.2, vertebrate physiology 8.8, immunology 8.2, other and unspecified 3.8)

9.5% (a) Mathematical sciences (108.3 as follows: mathematical statistics 67.2, mathematics 22.7, probability 6.2, computation 3.8, actuarial science 3.3, biomathematics 3.2, other 1.9)

6.3% (d) Public health (72.0 as follows: vital statistics 36.0, preventive medicine 15.3, hygiene 6.2, forensic medicine 3.7, other and unspecified 10.8)

3.4% (d) Human biology (39.0 as follows: psychology 24.2, population research 7.5, anthropology 4.3, unspecified 3.0)

3.1% (c) Clinical medicine (35.3 as follows: internal medicine 13.5, hospital statistics 3.8, other and unspecified 18.0)

2.5% (e) Social sciences and other (28.0 as follows: economics 19.0, education 3.8, other 5.2)

2.3% (e) Physical sciences (26.0 as follows: chemistry 8.3, engineering 8.0, physics 5.2, other 4.5)

The listing of interests in the 1953 Directory may be compared with that given in the 1949 Directory and analyzed by Tukey. The interests in both years have been grouped here into five main fields, which are laid down in Council By-Law Number 7 and which are important in elections to the Council. The results are:

	1949		1953	
	No.	%	No.	%
(a) Mathematics and Statistics	230	27.4	479	41.9
(b) General and Applied Biology	276	32.9	340	29.8
(c) Medical Sciences	136	16.2	158	13.8
(d) Human Biology and Public Health	118	14.1	111	9.7
(e) Other	78	9.3	54	4.7
Totals	838	99.9	1142	99.9

Direct comparisons of the number of members in the two years is biassed by 62 omissions in 1949 of members who listed no interests and only two omissions in 1953. The most striking change is that the mathematical and statistical interests now form the largest single group in the Society, although still outnumbered by the biological disciplines (b–d).

The figures for 1953 may be broken down on a geographical basis as follows:

Area	No. of interests assigned to group					Total
	(a)	(b)	(c)	(d)	(e)	
Belgium	13.8	23.3	18.8	3.2	4.8	63.9
Denmark	7.5	2.0	3.0	0.0	0.5	13.0
France	21.8	17.0	2.7	8.0	0.5	50.0
Germany	10.5	3.3	5.7	1.8	0.7	22.0
Italy	9.7	22.0	14.2	5.2	0.0	51.1
Netherlands	10.0	3.3	0.0	3.3	0.3	16.9
Sweden	8.7	0.5	0.0	0.3	0.5	10.0
Switzerland	7.7	2.8	1.7	1.8	0.0	14.0
United Kingdom	65.0	40.7	22.0	6.3	4.0	138.0
India	8.2	1.3	1.5	0.8	1.2	13.0
Japan	13.0	18.7	2.5	0.0	3.8	38.0
Eastern North America	208.7	127.5	69.5	67.0	26.3	499.0
Western North America	39.3	38.2	6.7	6.5	3.3	94.0
Brazil	5.0	2.0	1.3	0.0	0.7	9.0
Australasia	24.5	4.7	5.5	1.0	2.3	38.0
Other	25.3	32.7	3.3	5.7	5.0	72.0
Total	478.7	340.0	158.4	110.9	53.9	1141.9

In the same directory, the educational background of the membership is indicated by the highest reported degree.

Academic Degree	Frequency	
	No.	Percent
Medical (M.D., M.B., D.P.H.)	110	9.6
Other Doctorates	523	45.7
Masters	176	15.4
Bachelors	156	13.6
Other and not given	179	15.6
Total	1144	99.9

The information on type of employer has also been assembled. In cases where a member had two employers, each was given a weight of 0.5.

Type of Employer	Frequency	
	No.	Percent
Educational Institutions		
Professors and heads of departments	199.5	17.4
All others, from students to associate professors	292.5	25.6
Governmental Organizations		
Agricultural	136.5	11.9
Medical	107	9.4
Other	120.5	10.5
Non-profit Organizations, including research institutions, foundations and hospitals	61.5	5.4
Private Industry	171	14.9
Self-employed	23	2.0
Other	32.5	2.8
Total	1144	99.9

Australasian Region. The Regional Meeting was held in Canberra on 13 January 1954, in collaboration with Section D of the Australian and New Zealand Association for the Advancement of Science. Papers were presented by H. O. Lancaster on Significance levels in discrete distributions; by J. Biggers and P. Claringbold on The angular transformation in quantal analysis; by P. A. P. Moran on Dilution Assay of viruses; and by F. E. Binet and H. T. Clifford on A quantitative study of an assumed hybrid swarm between two Eucalyptus species. The meeting was well attended and the papers were followed by stimulating discussion. The Victorian and New South Wales Branches of the Region are quite active and have held monthly meetings throughout the year.

Italian Region. The IV Annual meeting of the Italian Region of The Biometric Society was held in Rome on March 28, 1954, in conjunction with the first meeting of the Associazione Genetisti Italiani, at the Consiglio Nazionale delle Ricerche.

There were two lectures, one in the morning by Prof. K. Mather of Birmingham University on "The Measurement of Genic Interaction" and one in the afternoon by Dr. C. A. B. Smith of University College, London, on "Funzioni Discriminanti e loro applicazione a problemi biometrici". Four papers were announced; of these one, by Drs. G. Barbensi and Nistri, on the Statistical study of electroencephalography,

was not given. The following papers were read at the meeting: L. Cavalli-Sforza, on the Calculation of chi square with small expected frequencies (a method based on the normalization of the Poisson distribution); R. Scossiroli on the Determination of heritability of body weight in the course of development, an experimental and statistical study on White Leghorn chickens; P. Dassat, on The possibility of anticipating and making more accurate the progeny test in its application to problems of milk production.

The main activities of the Italian Region for 1954–55 are dedicated to the organization, with I.U.B.S. support, of a summer course in Biometry, which is planned for early September, 1955. The course is expected to last just under three weeks and is designed to familiarise students who have some background knowledge of statistics with the use of more advanced methods. It should be open to graduates in medicine, biology and agriculture.

ENAR. The seventh annual meeting of the Region was held at the University of Florida in Gainesville on September 6 and 7, 1954. A morning session of contributed papers on September 6 featured papers by C. I. Bliss and A. R. G. Owen on Estimating k for a series of negative binomial distributions, and by W. D. Hanson and A. E. Willson on An adaptation of the method of paired comparisons to measure taste preference for a specific attribute. The annual business meeting which followed elected the following regional officers for 1955: President—D. B. Duncan, Secretary-Treasurer—A. M. Dutton, Regional Committee—(new members) S. Lee Crump and J. A. Rigney. Two joint programs on the afternoons of September 6 and 7 attracted a good attendance. The first, with the American Society of Plant Physiologists and the Botanical Society of America, concerned experimental topics and design, with invitational papers by E. J. Koch on The role and application of design principles in botanical research, by F. P. Steward and F. M. Shantz on The experimental control of variables and variability in the growth of tissue cultures, by S. Aranoff on Sampling problems in tracer work, and by H. Hopp on The use of discriminant functions in botanical research. The American Society for Horticultural Research co-sponsored the second joint program on The use of variance components in horticultural research, with papers by W. D. Hanson, by E. F. Shultz, by R. H. Sharpe and C. H. Van Middelem, and by D. D. Mason and R. J. Monroe.

On September 10–14, 1954, the Region met again, in Montreal, Canada, where it sponsored seven sessions jointly with the Biometrics Section of the American Statistical Association and in part with the Institute of Mathematical Statistics. The opening session concerned the Theory of Sampling Fish Populations, with papers by D. G. Chapman

on combined estimation methods and by D. S. Robson on tag-recapture procedures. Later programs considered Problems Associated with Mass Therapeutic Trials, with papers by D. J. Thompson on the estimation of the population group subject to benefit, and by D. D. Rutstein on a cooperative rheumatic fever study; Multiple Comparison and Multiple Decision Procedures, the first being discussed by J. Cornfield and the second by R. Beckhofer; Non-parametric Estimation of Survivorship, with papers on the survivorship of red blood cells by M. A. Schneiderman and on non-parametric estimation by Paul Meier; Analysis of Count Data (two sessions) with contributions by W. G. Cochran on some limitations of chi-square, by A. W. Kimball on short-cut formulas for the exact partitioning of chi-square, by J. Berkson on estimation by the principle of least squares and by maximum likelihood, by J. W. Tukey and F. J. Anscombe on transforming counted observations, and by N. Mantel on the interactions of certain contingency tables. In the closing session on Biological cycles D. A. MacLulich considered population dynamics from the standpoint of the biologist and Mark Kac the statistical problems and techniques in population cycle analysis.

Meeting of The Biometric Society (ENAR) Chapel Hill, North Carolina

The Biometric Society (ENAR) will meet jointly with the Institute of Mathematical Statistics at Chapel Hill, North Carolina April 22 and 23, 1955. Titles and abstracts for contributed papers for The Biometric Society should be sent to Dr. Bernard G. Greenberg, Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina not later than March 20, 1954.

BOYD HARSHBARGER, *Chairman*
Spring Meetings of the ENAR

BIOMETRICS

VOLUME 10: 1954
Numbers 1–4

Published Quarterly by
THE BIOMETRIC SOCIETY
NEW HAVEN, CONN.
1954

STATEMENT REQUIRED BY THE ACT OF AUGUST 24, 1912, AS AMENDED BY THE ACTS OF MARCH 3, 1933, AND JULY 2, 1946 (Title 39, United States Code, Section 233) SHOWING THE OWNERSHIP, MANAGEMENT, AND CIRCULATION OF

BIOMETRICS published Quarterly at New Haven, Connecticut for January-December, 1954

1. The names and addresses of the publisher, editor, managing editor, and business managers are:

Name	*Address*
Publisher, The Biometric Society	52 Hillhouse Ave., New Haven, Conn.
Editor, Prof. Gertrude Cox	Box 5457, College Station, Raleigh, North Carolina
Managing Editor, Mrs. Sarah P. Carroll	Box 5457, College Station, Raleigh, North Carolina
Business Manager, None.	

2. The owner is: (if owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual member, must be given.)

Name	*Address*
The Biometric Society	52 Hillhouse Ave., New Haven, Conn.

3. The known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages, or other securities are: (If there are none, so state.)

Name	*Address*
None	

4. Paragraphs 2 and 3 include, in cases where the stockholder or security holder appears upon the books of the company as trustee or in any other fiduciary relation. the name of the person or corporation for whom such trustee is acting; also the statements in the two paragraphs show the affiant's full knowledge and belief as to the circumstances and conditions under which stockholders and security holders who do not appear upon the books of the company as trustees, hold stock and securities in a capacity other than that of a bona fide owner.

5. The average number of copies of each issue of this publication sold or distributed, through the mails or otherwise, to paid subscribers during the 12 months preceding the date shown above was: (This information is required from daily, weekly, semiweekly, and triweekly newspapers only.).....

Sarah P. Carroll, Managing Editor

Sworn to and subscribed before me this 29th day of September, 1954.

[SEAL]

Katherine C. McKimmon
(My commission expires April 17, 1955)

BIOMETRICS

Contents of Volume X

ABSTRACTS	160, 302, 569
NOTES	192
QUERIES	155, 298, 407, 562
THE BIOMETRIC SOCIETY	189, 315, 414, 571

ABELSON, ROBERT M. and RALPH A. BRADLEY, A 2 × 2 Factorial with Paired Comparisons 487

ABELSON, ROBERT M. and RALPH A. BRADLEY, *Abstract* #273 303

ALMOND, JOYCE, A Note on the χ^2 Test Applied to Epidemic Chains 459

ANDERSEN, S. L. (*See* BOX, G. E. P.)

ANSCOMBE, F. J., Fixed-Sample-Size Analysis of Sequential Observations 89

BARBENSI, G., *Abstract* #240 161

BARGETON, D., *Abstract* #297 313

BAUER, R. K., *Abstract* #293 312

BENTZON, MICHAEL WEIS (*See* MEYER, SVEN NISSEN)

BENTZON, M. W., *Abstract* #267 185

BERKSON, JOSEPH (*see* DEWS, PETER B.)

BERKSON, J., *Abstract* #259 178

BIGGERS, J. D. and P. J. CLARINGBOLD, *Abstract* #301 569

BINET, F. E. and H. T. CLIFFORD, *Abstract* #300 568

BLISS, C. I., *Abstract* #241 162

BOX, G. E. P., The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples 16

BOX, G. E. P., and S. L. ANDERSEN, Abstract #272 303

BRADLEY, RALPH ALLAN, Incomplete Block Rank Analysis: On the Appropriateness of the Model for a Method of Paired Comparisons 375

BRADLEY, RALPH A. (*See* ABELSON, ROBERT M.)

BROSS, IRWIN, A confidence Interval for a Percentage Increase 245

BROSS, IRWIN, Misclassification in 2 × 2 Tables 478

CALVIN, LYLE D., Doubly Balanced Incomplete Block Designs for Experiments in which the Treatment Effects are Correlated 61

CLARINGBOLD, P. J. (*See* BIGGERS, J. D.)

CLIFFORD, H. T. (*See* BINET, F. E.)

COCHRAN, W. G., The Combination of Estimates from Different Experiments 101

COCHRAN, W. G., Some Methods for Strengthening the Common χ^2 Tests 417

DARMOIS, G., *Abstract* #261 180

DEWS, PETER B., GEORGE M. HIGGINS and JOSEPH BERKSON, Error of the Determination of the Eosinophil Count in Peritoneal Fluid of the Rat 221

DOERING, H., *Abstract* #294 312

FAVERGE, J. M., *Abstract* #270 302

FEDERER, W. T. and C. S. SCHLOTTFELDT, The Use of Covariance to Control Gradients in Experiments 282

FINNEY, D. J., *Abstract* #258 177

FISHER, R. A., The Analysis of Variance with Various Binomial Transformations 130

FISHER, SIR RONALD, *Abstract* #243 164

FRÉCHET, MAURICE, *Abstract* #260 179

FREE, S. M., *Abstract* #278 305
FREEMAN, G. H. (*See* HOBLYN, T. N.)
GEBELEIN, H., *Abstract* #284 308
GEIDEL, H., *Abstract* #285 308
GOODALL, D. W., *Abstract* #263 183
GRANDAGE, ARNOLD H. E., *Abstract* #277 305
GRAYBILL, FRANKLIN, Variance Heterogeneity in a Randomized Block Design 516
GRUNDY, P. M., D. H. REES and M. J. R. HEALY, Decision between Two Alternatives—How Many Experiments? 317
HAMAKER, H. C., *Abstract* #269 187
HAYMAN, B. I., The Analysis of Variance of Diallel Tables 235
HEALY, M. J. R. (*See* GRUNDY, P. M.)
HEALY, M. J., *Abstract* #249 168
HEMPHILL, F. M., Suggested Desk Calculator Operations for Computing Moments by the Row 152
HIGGINS, GEORGE M. (*See* DEWS, PETER B.)
HOBLYN, T. N., S. C. PEARCE and G. H. FREEMAN, Some Considerations in the Design of Successive Experiments in Fruit Plantations 503
HOBLYN, T. N. and S. C. Pearce, *Abstract* #257 176
HOLT, L. B., *Abstract* #254 173
HOPKINS, J. W., Incomplete Block Rank Analysis: Some Taste Test Results 391
HOPKINS, J. W., Some Observations on Sensitivity and Repeatability of Triad Taste Difference Tests 521
HOPKINS, J. W., *Abstract* #247 167
HOUSEHOLDER, A. S. (*See* KIMBALL, A. W.)
HUDDLESTON, HAROLD F., *Abstract* #281 306
HULL, FRED H., *Abstract* #283 307
HUNTER, J. S., *Abstract* #279 306
IPSEN, J., *Abstract* #253 172
KARREMAN, G., *Abstract* #266 185
KARREMAN, GEORGE, *Abstract* #282 307
KEITER, F., *Abstract* #287 309
KEULS, M., *Abstract* #248 167
KIMBALL, A. W. and A. S. HOUSEHOLDER, A Stochastic Model for the Selection of Macronuclear Units in Paramecium Growth 361
KIMBALL, A. W., Short-Cut Formulas for the Exact Partition of χ^2 in Contingency Tables 452
KNOWLES, E. A. G., *Abstract* #268 186
KODLIN, D. (*See* KRECH, U.)
KRECH, U. and D. KODLIN, *Abstract* #280 306
LANCASTER, H. O., *Abstract* #299 568
LEDERMANN, SULLY C., *Abstract* #296 313
LEGAY, J. M., *Abstract* #271 302
LEIN, A., *Abstract* #288 309
LI, C. C. and LOUIS SACKS, The Derivation of Joint Distribution and Correlation between Relatives by the Use of Stochastic Matrices 347
LOWRY, D. C., *Abstract* #245 165
LUSH, J. L., *Abstract* #246 166
MANDEL, JOHN, Chain Block Designs with Two-Way Elimination of Heterogeneity 251
MARTIN, L., *Abstract* #239 160
MARTIN, L., *Abstract* #251 171
MATHER, K., *Abstract* #244 164
MEYER, SVEN NISSEN and MICHAEL WEIS BENTZON, Further Studies on the Significance of Family Factors for the Response to BCG Vaccination: The Development of Local Vaccination Lesions and Their Relation to Allergy Production 195
MOORE, P. G., A Note on Truncated Poisson Distributions 402
MORAN, P. A., *Abstract* #298 568
MUENZNER, H., *Abstract* #289 310
NELDER, J. A., A Note on Missing-Plot Values 400
OWEN, A. R. G., *Abstract* #264 183

PANSE, V. G., *Abstract* #256 174
PARKER-RHODES, A. F., *Abstract* #262 181
PEARCE, S. C. (*See* HOBLYN, T. N.)
PETERSEN, R. G., *Abstract* #276 304
PRIGGE, R., *Abstract* #252 171
RAO, C. RADHAKRISHNA, Estimation of Relative Potency from Multiple Response Data 208
RAO, C. RADHAKRISHNA, *Abstract* #250 170
READ, D. R., The Design of Chemical Experiments 1
REES, D. H. (*See* GRUNDY, P. M.)
SACKS, LOUIS (*See* LI, C. C.)
SAMPFORD, M. R., The Estimation of Response-Time Distributions: III Truncation and Survival 531
SCHLOTTFELDT, C. S. (*See* FEDERER, W. T.)
SIECKMANN, W., *Abstract* #290 310
SMITH, C. A. B., *Abstract* #265 184
THOMPSON, W. A., JR., *Abstract* #275 304
TWEEDIE, M. C. K., *Abstract* #274 304
VAN DER REYDEN, D., Design and Analysis of Soil Insecticide Field Experiments 291
VESSEREAU, A., *Abstract* #242 162
VON GUÉRARD, H. W., *Abstract* #286 308
WALTER, E., *Abstract* #291 311
WELTE, E., *Abstract* #295 312
WETTE, R., *Abstract* #292 311
WORCESTER, JANE, How Many Organisms? 227
YATES, F., The Analysis of Experiments Containing Different Crop Rotations 324
YATES, F., *Abstract* #255 174
ZELEN, MARVIN, Analysis for Some Partially Balanced Incomplete Block Designs Having a Missing Block 273

BIOMETRICS

INDEX

Volumes 1–10

THE BIOMETRIC SOCIETY
NEW HAVEN, CONN.
1954

INTRODUCTION

BIOMETRICS INDEX, VOLUMES 1–10

This index is detailed and generously cross-indexed to facilitate its use by those who are relatively unacquainted with biometry. A few rules which have been followed are stated for the users' guidance.

First, only one reference to any one topic will be found for each paper, abstract, or query, except that a second reference is given when a long paper returns to the topic in question after at least ten pages. The user should therefore look forward for additional material. Also, he should look *backward*, because the index refers to the page on which the indexed topic is named (usually) and this is sometimes the conclusion of a discussion.

Second, index references sometimes lead to a page on which the index topic is not specifically mentioned. For example, the design mentioned in volume 1, page 48, toward the bottom of the first column, is a Latin square and is so indexed, though that term does not appear.

Third, modifying phrases such as "method of" have been ignored. This is not true of adjectives, e.g. "minimum".

Fourth, a few listings have an element of editorializing, as when the statement in volume 2, page 17, beginning at line 15 of the second column is indexed under "inference".

Fifth, abbreviations have been avoided, but a few occur.

Sixth, errata are not listed under similar terms such as corrigenda.

HORACE W. NORTON

Absorbing state, 9:285
Acceptance area, 6:63, 192; 8:185, 299; 9:25, and see rejection area, tests
computation, 6:62; 8:185
Accidents, 7:340; 8:275; 9:188, 266, 486; 10:245, and see Greenwood-Yule distribution
Action, see future action
Actuarial statistics, 1:4; 3:58, 129; 8:30, 90; 10:308, and see demography. Gompertz curve, life tables, medicine
Additivity, 2:24; 3:10, 23, 40, 50; 4:24, 254; 5:72, 84; 6:8; 7:14; 9:20, 290, 347, 389; 10:135, 145, 235, 562, and see analysis of variance, hypothesis, model, transformations
non-additivity, 3:35; 5:84
test of, 5:232; 10:413
Age distribution, 3:113; 5:19, 27, 217; 6:253; 7:156, 301, and see demography
Agronomy, 2:18, 36; 3:29, 42; 4:132, 171, 211, 234; 5:252; 7:39, 247; 8:386; 9:157, 296, 395, 498; 10:103, 118, 160, 162, 174, 304, 516, and see cereal chemistry, field experiments, field sampling, horticulture, long-term experiments, Mitscherlich's law, rotation experiments, selection of levels, soil
Aliases, 6:236; 10:5
separation of, 6:246
Alinement charts, see nomography
Allometry, 4:14; 9:48, and see growth
All-or-none response, see quantal response
Analgesia, 2:87; 8:218, and see pharmacology
Animal breeding, 1:27; 2:21, 56; 3:54; 4:269; 5:71, and see inbreeding, plant breeding
Animal experiments, 2:54, 3:69; 4:5, 100; 6:193, 319, 363; 8:169; 9:430; 10:290, 304, and see design of experiments, poultry, twins
use of identical twins, 8:187
Anthropology, 1:20; 2:50, 58, 97; 3:193; 4:247; 6:181; 7:122, and see demography, medicine
Antibiotics, 1:10, 40; 5:211, 317, 333; 6:242, 325, 347; 10:435
Approximate formulas, 2:96; 5:325; 7:266; 8:40, 298
Approximate significance levels, 2:74, 112; 3:28, 120; 4:263; 5:101, 325; 6:178, 191
Ascorbic acid, see vitamins
Association, see contingency
Attenuation, 2:103; 7:36
Average sample number, 6:65; 8:200; 9:263; 10:312
Bacteriology, 2:82; 3:175; 5:126, 163, 211, 277, 283, 289, 317, 333; 6:105, 120, 128, 169, 181, 290, 347; 7:120, 133, 271, 320, 334; 8:391; 9:185, 201, 260, 278, 320; 10:227, 423, 541, and see antibiotics, dilution series, extinction time, microbiological assay, sanitary engineering
Balancing, 8:170, and see paired comparisons, randomized blocks, selection
Bartlett's test, see homogeneity of variance, multivariate tests
Bayes' distribution, 10:308
Behrens-Fisher test, 3:64; 5:337; 6:192; 7:5, 19; 9:61
power of, 6:192
Beta function, incomplete, 6:275, 378; 7:6; 9:13, 28, 282, and see F test
Between-block error, see interblock error
Bias, 1:13, 49, 84; 2:1, 3, 18, 22, 41, 44, 54, 76, 97, 111; 3:26, 94; 4:265; 5:31, 85, 297; 6:13, 100, 155, 176; 7:50, 160, 289, 297, 310, 353; 8:93, 234, 283, 290; 9:1, 61, 75, 258, 532; 10:5, 32, 119, 141, 178, 223, 263, 305, 470, 481, 545, and see Sheppard's correction
Binomial distribution, 1:43, 44, 60, 76, 81, 84; 2:97; 3:42, 45, 166, 175; 4:255; 6:1, 107, 180, 299, 346, 392; 7:276, 328; 8:1, 87, 221, 268, 323, 391; 9:28, 282, 291, 364, 467; 10:92, 130, 178, 246, 425, 464, 480, 523, and see distributions, Galton's quincunx, multinomial distribution, quantal response, sign test

compound, 6:80
modified, 6:173
negative, 3:43; 5:162, 165; 6:60, 103, 193; 7:356, 403; 8:281, 287, 390; 9:176, 197, 354; 10:167, 304, 419, 430
truncated, 10:405
Bioassay, 1:10, 39, 40, 41, 51, 57, 63; 2:30, 34, 83, 96, 109; 3:60, 62, 69, 172, 173; 4:104, 122, 153, 197; 5:83, 85, 86, 87, 127, 211, 261, 273, 300, 317, 330, 335; 6:99, 174, 175, 176, 320, 322, 413, 424, 432; 7:121, 200, 295, 302, 323, 327; 8:13, 51, 120, 176, 177, 206, 237, 370; 9:260, 266, 290, 304, 390, 405, 428, 431, 471; 10:178, 185, 208, 305, 306, 310, 541, 555, and see biometry, dose response curve, extinction time, Karber's method, microbiological assay, pharmacology, quantal reponse, Reed-Muench method, selection of dosages, sensitivity tests, survival time, time response curve, toxicology, transformations
factorially designed, 1:62; 2:83
human subjects, 8:232
interpretation of, 5:273; 6:100
non-linear response, 5:282
Biochemistry, 4:52, 144, 202; 5:329; 6:122, 322, 327, 342, 347, 395, 435; 7:39, 133, 185, 331; 8:175; 9:431
Biological assay, see bioassay
Biological measurement, see bioassay
Biological variation, 2:101; 3:15; 4:49, 199, 219; 5:128
Biomathematics, see biometry
Biometrics, see biometry
Biometric Society, 3:184; 4:82, 145, 216, 273; 5:90, 180, 254, 341; 6:69, 195, 300, 438; 7:127, 224, 303, 436; 8:97, 180, 273, 392; 9:111, 271, 535; 10:189, 315, 414, 571.
Biometry, 1:41, 47; 2:5, 34, 48, 82, 122; 3:61, 66, 145, 172; 4:1, 14, 91, 139, 140, 141, 143, 144, 156, 197, 217; 5:24, 117, 142, 273, 323; 6:75, 85, 145, 183, 184, 250, 320, 322, 330, 342, 347, 395, 413; 7:123, 134, 180, 186, 227, 335, 353, 391, 435; 8:13, 91, 100, 174, 210, 271, 307, 389, 390; 9:47, 90, 117, 201, 212, 260, 275, 320, 461, 471, 522; 10:147, 172, 177, 179, 185, 200, 227, 302, 307, 308, 313, 347, 361, 412, and see allometry, bioassay, growth, hypotheses, Mitscherlich's law, models, path coefficients, Weber-Fechner law
Biophysics, see biometry
Birth rank, 5:87; 8:112; 9:265; 10:442
Birth rate, 6:252; 7:155; 8:272; 9:212, and see demography
Block size, 10:73, 506, and see plot size
Blood counts, 3:175; 6:282, 390; 7:171; 8:55; 10:167, 222
Blood groups, 1:20, 35; 6:174
Borders, 2:57; 6:203; 9:158; 10:175, 297, and see plot size
Brownian motion, 6:183; 9:460
Calculation, see computation
Cancer, 1:3; 3:123; 4:51, 139; 6:103; 8:3, 16, 90, 271, 320; 9:278; 10:313, 431, 540
Canonical analysis, 10:7, 14, 35
Carry-over effect, 6:315; 8:230; 9:430; 10:176, 505, and see Latin squares in time
Case histories, 5:76, 77
Causation, 1:3, 75, 86; 2:15; 10:309, and see correlation
Cell division, 3:181; 7:133; 10:361, 455
Censored data, 8:308; 10:533, and see truncation
Census, 1:11; 7:155, 257, and see demography
sample 1:30; 2:76; 7:300; 8:286
Cereal chemistry, 2:26
Chain blocks, 7:124, 125; 9:127, 141, 263; 10:251
generalized, 10:251
table of, 10:256
Change-over design, see cross-over design
Characteristic function, 7:277; 8:117; 9:355, and see cumulants, moments
Check plot, 1:13, 14; 2:36, 57; 3:44; 5:83; 10:291, 298, 300, and see standard variety
Chemistry, 1:27, 41, 50; 3:61; 4:202; 5:87; 6:228, 233, 295; 7:56, 102, 330; 8:386; 9:135, 259; 10:1, 19, 306, and see biochemistry, cereal chemistry,

chain blocks, chemotherapy, compositing, industrial research, mass action, quantitative analysis, random walk, sanitary engineering, spectrography
Chemotherapy, 1:4, 59; 7:171
Chi square, 1:10, 15, 28, 39, 43, 60, 61, 76, 81, 84, 85; 2:9, 17, 37, 47, 74, 96, 111; 3:18, 91, 98, 125; 4:80, 137; 5:83, 88, 169, 251, 337; 6:17, 29, 60, 133, 137, 148, 178, 182, 185, 218, 259, 291, 298, 308, 321, 361, 375; 7:4, 114, 222, 283; 8:4, 18, 61, 114, 221, 272, 276, 288, 322, 345, 380; 9:108, 176, 255, 322, 354, 367, 473; 10:223, 233, 381, 397, 417, 459, 484, and see analysis of variance, contingency, continuity correction, fourfold table, gamma distribution, goodness of fit, minimum chi square estimate
  analysis of, 5:140, 301; 6:178; 7:124; 9:259; 10:184, 417, 452
  computation, 1:11, 55, 85; 4:80, 7:124, 283; 8:380; 10:453
  inappropriate, 10:446
  interpretation of, 7:115; 10:417
  noncentral, 7:31; 10:419
  power of, 10:142, 417, and see goodness of fit
  small expectations, 6:17; 10:418
  table, 2:112
Choice, see selection
Chromosomes, 1:7; 2:78; 4:1
Classification, see grouping, non-numerical data
Class mark, 8:20; 10:532, and see histogram
Clinical experiments, 1:3; 2:87, 95; 6:328; 8:188, 206, 218, 232; 9:260; 10:312
Clinical statistics, 1:3; 2:47; 3:110; 5:76, 95; 6:103, 328: 7:126; 8:174, 387
Clinical tests, 1:3; 5:77, 78; 6:399; 7:185, 240, 379, and see critical point, diagnosis, placebo
Coding, 5:79; 6:35; 8:324; 9:478, and see class mark
Coefficient of determination, 1:86; 5:123, and see correlation
Coefficient of variation, 1:50, 67; 2:3, 78; 3:26; 4:22; 5:161; 6:49, 110, 118; 8:59, 92, 108, 177; 9:43; 10:99, 221, 283, 296
  of product, 8:64
Collaborative experiments, 6:322, and see comparisons of laboratories
Combination of data, 6:176; 7:121
Combination of estimates, 9:59; 10:101, 155
Combination of tests, see tests
Communications, 8:268; 9:286
Comparisons, see analysis of variance, degrees of freedom, orthogonal polynomials, tests
  dummy, 6:295
  individual, 4:61, 106; 5:236
  interblock, 2:55
  intrablock, 2:55
  method of, 2:74
  non-orthogonal, 4:215, 240
  of laboratories, 1:10; 6:290; 8:237; 10:103, 273
  of means, 5:84, 99, 251
  orthogonal, 2:61, 84; 3:7, 74; 4:71, 106, 133, 173, 215; 5:286; 8:376; 9:387 paired, 1:56, 63, 80; 2:30, 38, 50, 95; 3:25, 89, 92, 119; 4:211; 7:125, 312; 8:87, 95, 189; 9:32, 43, 109; 10:303, 375, 487, and see half-leaf method
  selection of, see selection
Complement fixation, 4:200
Components, see correlation, covariance, variance
Composite design, 10:6, 33, 306
Compositing, 7:85
Computation, 1:69, 78; 2:15, 54, 55, 61, 84, 111; 3:62, 194; 5:55, 72, 289; 6:10, 221, 286, 305, 328, 360, 414; 7:200, 252, 337; 8:17, 36, 156, 179, 249, 325; 9:180, 291, 322, 358, 377; 10:31, 112, 138, 152, 172, 368, 421, 499, 551, and see coding, estimation, Gauss multipliers, graphics, iteration, Lagrange multipliers, matrices, method of moments, method of steepest descent, Newton's method, nomography, recurrence formulas, service, simultaneous equations, slide rule
  accuracy of, 3:194; 8:179; 9:183, 307, 358; 10:432

approximate, 1:69
Conclusions, 1:8, 85; 2:29, 38, 48; 3:20; 5:96; 6:371; 7:101; 8:113; 9:209, 532; 10:308, and see future action, inference, tests
Concomitant observations, 2:28; 6:303; 8:169, and see analysis of variance, statistical control
Confidence limits, 1:8, 38, 57, 59, 76, 2:4, 9, 15, 63, 85, 96, 104, 111, 120; 3:14, 22, 86, 167; 4:185, 214, 237; 5:177, 208, 240, 252, 266, 283, 311, 333, 335; 6:14, 80, 101, 114, 175, 185, 194, 321, 323, 327, 394, 425, 435; 7:51, 94, 164, 200, 259, 301; 8:27, 40, 121, 179, 252, 287, 375, 391; 9:209, 313, 322; 10:94, 157, 168, 171, 182, 209, 304, 305, 309, 479, 555, and see tests
approximate, 1:59; 3:86; 6:137; 7:5
exact, 7:11; 8:375; 10:73
for correlation, 9:110
for function, 8:348
for missing value, 10:401
for percentage increase, 10:245
for proportion, 8:87; 10:171
for ratio, 5:336; 8:86
for regression, 9:205
for regression coefficient, 8:269
for variance components, 6:136; 7:5, 80
interpretation of, 2:107; 5:297; 7:198
multivariate, 7:53
not fiducial, 5:336
Confounding, 1:11, 31, 63; 2:55, 85; 3:69, 176; 4:54, 110, 240; 5:74, 145; 6:228, 236, 241, 329; 8:75, 123; 9:158, 298, 388, 434; 10:326, and see analysis of variance, design of experiments, factorial experiments
complete, 2:43; 9:299; 10:72
partial, 9:3, 158
table of systems, 4:61
Consultation, see service
Contagious distribution, 1:44; 2:76; 3:193; 7:405; 9:354; 10:167, 421, and see Polya distribution
Neyman's, 2:76; 6:103; 7:361; 9:184, 354
Contingency, 1:11, 56, 84; 2:16, 17, 47; 3:123; 6:178, 299; 7:283; 8:2, 61, 380; 9:108; 10:417, 452, and see chi square, correlation, fourfold table
Continuity correction, 1:11, 56, 84; 2:95; 3:46, 125; 6:299; 7:222; 8:222; 9:110; 10:137, 390, 420
inappropriate, 1:84; 5:339; 10:443
Contrasts, see comparisons
Control chart, 1:10, 40, 42, 74; 3:172; 4:125; 6:60, and see statistical control of quality
Control plot, see check plot
Correlation, 1:15, 31, 71, 75, 86; 2:27, 47, 55, 56, 67, 101; 3:45, 54, 98; 4:21, 40; 5:30, 88, 119, 172, 228; 6:119, 259; 7:34, 60, 140, 356; 8:3, 295, 330; 9:91, 108, 496; 10:184, and see attenuation, causation, coefficient of determination, contingency, correlated errors, covariance, regression, sampling, scatter diagram, z transformation
autocorrelation, 9:279, and see serial correlation
between relatives, see statistical genetics
biological, 2:48
biserial, 4:137; 6:400
canonical, 10:306
combination of, 1:86
components of, 9:352; 10:355
genetic, 9:251, 339; 10:355
intraclass, 2:57; 3:54; 4:269; 9:266
maximum, 3:105
mean with variance, 2:85; 6:318, 435; 10:117, 196
multiple, 2:27
non-normal data, 2:73
organic, 9:47; 10:180
partial, 2:27; 3:123; 4:163; 6:344; 8:11, 240
part with whole, 3:65; 4:160
rank, 2:69; 3:138; 9:108
ratio, 1:75; 2:56, and see analysis of variance
serial, 7:315, and see time series
spatial, 7:300, 9:420
spurious, 2:50; 3:65; 5:76; 7:366
tetrachoric, 3:126, and see chi square, fourfold table
within blocks, 10:61

Counting, 8:105; 10:221, 446, and see bacteriology, blood counts
Course of action, see future action
Courses in statistics, see statistics curricula
Covariance, 1:15, 20, 38; 2:15, 21, 43, 78; 3:97; 4:30, 195; 6:29; 7:34; 9:12, 222, 432; 10:272, 282, 307, 340, and see concomitant observations, correlation, dummy variates, error regression, matrices, regression, statistical control, variance
  adjustment, 3:34; 8:237; 10:176
  analysis of, 1:77; 2:21, 27, 53; 3:60, 65; 4:48, 153, 240; 5:72, 178, 279, 289; 6:15, 164, 184, 297, 303; 7:37, 124, 191, 313; 8:170, 387; 9:342, 437, 531; 10:285, 297, 504
    computation, 4:158; 5:72; 7:124; 9:432
    multiple, 5:117; 6:293
    multivariate, 6:381
  components of, 9:226, 341
  matrix, 3:104, 166; 4:178; 6:181, 213, 262, 376; 7:137, 247; 8:5, 24, 156, 335; 9:49, 207, 495; 10:11, 518, 534, 547
  multiple, 3:35; 6:293
  of estimates, 9:508
  of estimates of variance components, 7:5
  of products, 6:411
  of ratios, 6:411
Critical point, 6:403; 8:36
  approximate, 8:39
Critical ratio, 6:298
Critical region, see rejection area
Crossing over, 4:1; 6:182, and see linkage
Cross-over design, 1:63; 3:70; 6:193, 301, 312; 7:317; 9:430, 431, and see Latin square in time
Cumulants, 7:5, 140; 8:112; 9:11; 10:429, and see characteristic function, moments generating function, 10:543
Curve fitting, 2:82; 3:56; 4:197; 5:207; 7:236, 337; 10:303, and see estimation, fitting regression line, growth, orthogonal polynomials
Cut, significant, 7:50
Cycles, 1:7; 4:230; 5:38, 121, 219; 6:251; 9:280, 429, and see time series
Cytology, see cell division, crossing over
Death rate, see mortality
Decision, 2:16; 4:215; 5:100; 8:196, and see conclusions, tests
Defining contrast, 6:238
Degrees of freedom, 1:19, 37, 56, 60, 61, 85; 2:28, 35, 85, 111, 120, 121; 3:54, 64, 99; 4:238, 264; 5:232; 7:20, 317; 8:247; 9:60, 130, 502; 10:25, 110, 125, 303, 328, and see comparisons
  adjusted, 10:111
  effective, 7:8
  equivalent, 9:63
  fractional, 2:112; 10:410
Demography, 1:20; 2:58; 3:193; 7:155, 257; 8:272, and see actuarial statistics, age distribution, anthropology, birth rate, blood groups, census, death rate, medicine, public health, sex ratio
Dependent variable, see regression
Design of experiments, 1:11, 13, 16, 31, 42, 47, 49; 2:18, 26, 29, 36, 57, 77, 86, 109, 113, 123; 3:20, 175; 4:100, 123, 171, 213, 215, 240, 269, 277; 5:70, 169, 178, 252, 279; 6:110, 118, 127, 167, 201, 228, 301, 303, 376, 390, 424; 7:6, 51, 61, 85, 204, 252, 273, 312; 8:169, 206; 9:31, 39, 110, 260, 276, 290, 304, 428, 502, 522; 10:1, 89, 174, 176, 177, 183, 186, 187, 291, 308, 309, 312, 503, and see aliases, balancing, block size, chain blocks, components of variance, composite design, concomitant observations, confounding, cross-over design, defining contrast, dummy treatments, factorial experiments, field experiments, fractional replication, half-leaf method, incomplete blocks, interblock error, interblock information, intrablock error, lattices, long-term experiments, multifactor design, non-orthogonal design, optional stopping, organolepsis, orthogonal arrays, orthogonal squares, paired comparisons, parallelogram design, perennials, placebo, plot size, precision, randomization, ran-

domized blocks, replication, risk function, rotating factors, sampling, sequential experiments, split plots, staggered design, standard variety, systematic design, tree crops, triangle test, Youden square
for locating optimum, 10:3, 16, 177, 306, 323
Determinants, see matrices
Deviation, see variance
Diagnosis, 1:3; 3:110; 4:100; 5:76, 77; 6:172, 173, 282, 399; 7:241; 10:478, and see clinical tests, medicine
Differential equations, 4:207; 5:323; 6:356; 9:213, 286, 462
Digitalis, 1:10, 57; 3:65; 6:322
Dilution curve, 6:170
Dilution series, 2:82; 4:101; 5:128, 325; 6:105, 128, 181; 7:121; 9:333; 10:227, and see bacteriology, virology
Discriminant function, 3:104, 180; 5:78, 80; 6:184, 303; 7:122; 8:387; 9:34, 117; 10:170, 184, 312, and see multivariate analysis, scores
Disproportionate subclasses, 1:11; 2:21; 6:179; 7:11, 118, 318; 9:226, 253, 264, and see analysis of variance, fitting constants, method of expected subclass numbers, unequal subclasses
computation, 2:35
Distribution, 3:174, 4:218; 6:103, 110, 128, 174; 8:13, 93, 307, and see accidents, approximate significance levels, Bayes, binomial, characteristic function, chi square, contagious, cumulants, *D* statistic, Edgeworth series, exponential, gamma, Gompertz curve, goodness of fit, Gram-Charlier series, Greenwood-Yule, histogram, hypergeometric, intercept, J-shaped, Kapteyn's, kurtosis, logarithmic, logarithmic normal, moments, normal, Pearsonian, percentile, Poisson, Polya, populations, quantile, quartile, rectangular, sampling, skewness, studentized deviate, Thomas, U-shaped
approximate, 2:111; 6:259; 8:94; 9:60
asymptotic, 7:20; 8:288, and see normal distribution
bivariate, 2:97; 7:413; 8:332
compound, 7:119; 9:86
resolution of, 9:86
conditional, 3:164; 6:194; 7:22; 8:300; 10:397, 486, and see conditional probability
contagious, see contagious distribution
cumulative, 2:74
discontinuous, 8:117
discrete, 6:103, 128; 7:365; 9:11
empirical, 6:192
exact, 2:111; 5:81; 6:19, 178, 259, 285; 7:125; 8:112; 10:418
finite, see populations
fitting, see estimation
frequency, 1:73; 8:33
infectious, see contagious
joint, 6:28
marginal, 8:312; 9:47
mixing, 6:194
multivariate, 3:14; 5:78, 80, 242; 6:175, 259, 372; 7:181; 9:47; 10:180, 518
Neyman's, see contagious
normality of, 5:81
nuisance, 8:313
of chi square, 10:460
of differences, 8:85
of estimates, 6:415
of estimates of variance components, 2:110; 7:5
of events, 6:171; 7:134
of *F*, 3:24
of gaps, 5:109
of rank correlation, 2:69
of rank totals, 1:81; 2:68; 8:33
of scores, 2:68; 8:117
of *t*, 3:24
of weighted mean, 9:448
of *z*, 3:99
posterior, 6:190
prior, 6:194
radial, 9:460
spatial, 1:44; 2:5
truncated, 6:186; 7:120; 8:15
Dominance, 4:254; 6:154, 353; 7:18; 9:90, 349; 10:165, 235, 357
Doolittle method, 6:179, 381; 7:11, and see matrices
Dose response curve, 1:10, 40, 41, 50, 57; 2:34, 82, 85, 109; 3:74; 4:125, 197;

5:83, 85, 262, 281, 321, 330; 6:175, 326, 345; 7:122, 201, 335; 8:219, 239, 250; 9:305, 320; 10:173, 185, 227, and see time response curve
table, 9:324
Drosophila, 1:7, 55; 2:9, 78, 110; 4:9; 6:104, 214; 7:336; 10:239
D statistic, 6:175, and see multivariate analysis
Dummy treatments, 9:387, 403
Dummy variates, 2:43; 3:34; 4:241; 6:64, 311, and see covariance, missing values
Duo-trio test, 9:43, and see triangle test
Duplication, see replication
Ecology, 1:7, 11, 44, 47, 76; 2:1, 76; 3:59, 101, 145; 5:14, 28, 115, 162, 213; 6:59, 250, 353; 7:300; 8:89, 177, 286, 386; 9:117, 176, 275, 369; 10:183, and see entomology, fish, genetics, natural selection, peck order, populations, quadrat sampling
Economics, 1:67; 3:110; 6:256; 7:36, 100; 8:186; 9:505, and see cost of sampling Gompertz curve, identification, stochastic processes, structure, time series
ED 50, 1:39, 41, 57, 59; 3:172; 4:200; 5:85, 127; 6:229, 329, 413, 432; 7:126, 295, 302, 339; 8:29, 51, 120, 249; 9:320; 10:147
tables, 8:53, 253
Edgeworth series, 8:117
Efficiency, 1:11, 12, 14, 42, 44, 78, 81; 2:2, 29, 31, 34, 36, 54, 57, 78, 119, 123; 3:22, 25, 42, 75, 175; 4:74, 112, 122, 191; 5:5, 156, 166, 179, 199, 208, 288; 6:15, 45, 61, 189, 192, 202, 221, 247, 282, 302, 392; 7:10, 50, 101, 113, 158, 169, 217, 223, 247, 292, 302; 8:30, 92, 126, 170, 173, 204, 236, 279, 334; 9:103, 179, 197, 264, 304, 333, 447, 471, 476, 487; 10:33, 64, 286, 311, 317, 335, 482, 501, 509, 538, 553, and see estimation, maximum likelihood, nonparametric tests, sufficient statistics
asymptotic, 6:223; 7:338; 9:60
table, 7:219, 293; 10:539
Empirical formulas, 5:319; 9:458
Endocrinology, 1:11, 63; 2:85; 3:70, 99, 173; 4:122, 269; 5:49, 71, 176, 262, 278; 6:98, 363; 7:122, 323; 8:206; 9:431, 467; 10:313
Entomology, 1:12, 30, 47, 51; 2:1, 30, 57, 76; 3:44, 99, 193; 4:171; 5:83, 165; 6:257; 7:121; 8:172, 277, 383; 9:117, 178, 290, 365, 425; 10:119, 142, 302, 455, 505, 540, 562, and see Drosophila, toxicology
Epidemiology, 1:44; 3:173; 5:243; 7:320; 8:275; 9:285; 10:459, and see contagious distribution, infection interval, medicine, Polya, public health, rumors
Equiprobable ellipse, 4:28
Ergodicity, 9:284
Errata, 3:54, 138, 140; 4:246; 6:33, 168, 258; 7:222, 274; 9:106, 335, 534
Errors, theory of, 2:103; 3:23; 5:301; 7:33, 309; 9:92, 312, and see analysis of variance, attenuation, extrapolation, interpolation, least squares, maximum likelihood, models, rejection of data
correlated, 2:18; 7:58, 309; 10:61, 516, and see *F* test
correlation, 2:53; 3:32; 4:163
distribution, 3:26, and see normal distribution
gross, 3:27; 5:72, 288; 9:75; 10:425
heterogeneity of, 1:11; 2:27, 113; 3:28, 37; 6:14, 100; 10:516
interblock, 1:58; 4:60; 5:151; 6:25
intrablock, 1:58; 4:60; 5:154; 6:25; 8:75; 9:422
matrix, 3:377
mean square, 10:482, and see variance of classification, 10:478
of estimation, 5:168; 7:140
propagation of, 8:388
regression, 2:15, 43, 54; 4:159; 6:166, 297, and see analysis of covariance heterogeneity of, 1:11
term, proper, 1:18, 24, 27, 41, 43, 50, 62, 70; 3:25, 40, 91; 4:132, 268; 5:69, 72, 178; 6:166, 289, 434; 7:14, 70, 168; 8:169, 243, 385; 9:170, 312, 337; 10:236, 330, 407
Erythrocytes, 6:282; 7:171, 333, and see blood counts, hematology

Estimation, 1:28, 77; 2:96, 102; 3:3, 25; 4:6, 27, 58, 137, 219, 240, 248, 256; 5:33, 68, 84, 116, 165, 288, 306, 320; 6:25, 176, 181, 182, 186, 190, 221, 227, 251, 315, 390, 401, 413, 432; 7:120, 135, 240, 275, 295; 8:3, 13, 51, 183, 269, 278, 307; 9:33, 59, 75, 90, 121, 201, 262, 264, 291, 320, 375, 462, 485, 489; 10:94, 178, 180, 181, 184, 293, 402, 469, 481, 531, 557, and see bias, combination of estimates, efficiency, fitting constants, fitting regression line, game theory, information, least squares, maximum likelihood, method of moments, minimax, minimum chi square, minimum unlikelihood, minimum variance, populations, residual effects, scores, Sheppard's correction, small samples, sufficient statistics, theory of errors
graphical, 8:179
inefficient, 6:109
of components of covariance, 9:226
of derivatives, 10:5
of function, 8:27, 348
of maximum, see estimation of optimum
of median, 9:75
of optimum, 2:63; 4:171; 10:1
of potency, 3:86; 4:106, 124, 200; 5:266, 281, 309, 320; 6:99, 322, 345, 432; 8:19, 179, 249, 372; 9:313, 328; 10:209
of variance, 3:11, 97; 5:209; 6:60, 187; 7:66, 202; 8:77, 93, 237; 9:60, 462, 487; 10:238, 401
of variance components, 2:110; 4:229; 5:191; 6:179, 187; 7:3, 29, 70, 75, 85; 9:226
sequential, 6:193
unbiased, 2:97; 3:9; 6:286, 292; 7:4, 156, 279; 8:91, 269
Eulerian distribution, see gamma distribution
Evolution, 1:7, 27; 4:143, 218, 250; 8:177, 268; 9:117, and see ecology, genetics, natural selection, populations, taxonomy
Expectation, mathematical, 1:18; 2:8, 111; 3:119, 163; 4:64, 257; 5:86, 109, 308; 6:28, 60, 284; 7:3, 89, 140, 269, 285; 8:112, 239, 288; 9:56, 91, 197, 210, 232; 10:57, 143, 159, 182, 199, 330, 343, 381, 400, 427, 448, 459, 480, and see mean, moments
Expected subclass numbers, method of, 2:10; 7:11
Experimental design, see design of experiments
Exponential curves, 3:56; 6:395; 7:126, 247, and see asymptotic regression
Exponential distribution, 8:16
Extinction time, 5:127
Extrapolation, 5:229; 7:253; 10:4, 474, and see estimation, interpolation, theory of errors
Extreme deviate, 5:288; 9:13, 74, and see order statistics, tests
studentized, 5:107
Factor analysis, 7:61; 8:172; 10:22, 181, 183, and see multivariate analysis
Factorial analysis, see orthogonal comparisons
Factorial experiments, 1:10, 11, 31, 62; 2:26, 54, 77; 3:33, 40, 69, 175; 4:54; 5:145; 6:15, 79, 203, 228, 319, 362; 7:323; 8:1, 237, 387; 9:2, 169, 256, 259, 295, 304, 428, 467, 476, 502; 10:4, 303, 487, and see fractional replication, pseudofactorial
confounded, 2:43; 8:123; 9:158, 428
incomplete, 9:428
partially, see dummy treatments
Fatigue, 2:13; 7:345; 9:29, 44, 306; 10:61
Faulty analysis, 2:29; 3:101; 5:72; 8:3, 221
Fiducial limits, see confidence limits
Field experiments, 1:13, 26, 47; 2:18, 36, 44, 54; 3:33; 4:54, 109, 171; 5:1, 68, 144, 252; 6:201, 302, 303, 317; 7:21, 39, 167, 261, 300; 8:271; 9:157, 186, 395, 412; 10:291, 324, and see agronomy, borders, check plot, correlated errors, design of experiments, irrigation, long-term experiments, Mitscherlich's law, perennials, plot size, rotation experiments, soil, standard variety, tree crops, uniformity data
correlation with laboratory experiments, 1:42

optimum block shape, 5:145, 156
Field sampling, 1:29, 43; 2:1; 3:20, 41; 4:234; 5:190, and see forestry, soil
Fish, 1:36, 50, 73; 2:73; 3:145; 5:14, 27, 115, 213; 6:59; 7:171, 301; 8:286, 391, and see ecology, Lee's phenomenon, optimum catch, population management
scales, 1:73, 76; 5:15, 31, 217
Fitting constants, 1:37; 2:10, 22, 35, 43, 94; 4:245; 7:11, 118, 318; 8:3; 9:264, 293, 482; 10:236, 324, and see disproportionate subclasses, dummy variates, least squares, missing values
Fitting distribution, see estimation
Fitting regression line, 1:36; 2:28, 61; 3:162, 165; 5:320; 6:119, 180, 363; 7:23, and see covariance, least squares, regression
both variables subject to error, 2:101; 4:49; 5:87, 88, 207; 7:33, 106; 8:269; 10:167
Food technology, 2:64, 76; 3:104; 4:80; 5:190; 6:1, 295; 8:78; 9:39, 295; 10:24, 388, and see cereal chemistry, organolepsis
Forensic statistics, 6:435
Forestry, 1:16, 29, 67; 3:101; 5:174; 10:162, and see wood decay
Fourfold table, 1:55, 84; 2:47, 95; 3:123; 5:251; 8:220; 10:420, 441, 453, 478, and see chi square, contingency, continuity correction
F test, 1:19, 25, 26, 49, 70, 80; 2:14, 16, 27, 37, 38, 44, 56, 62, 74, 113; 3:14, 24, 54, 94, 96, 98, 137, 195; 4:268; 5:75, 102, 242, 250, 265; 6:130, 178; 7:6, 19, 51, 125, 168, 299; 8:125; 9:20, 147, 337, 414, 473, 529; 10:70, 108, 167, 196, 261, 287, 408, 440, 516, 562, and see analysis of variance, incomplete beta function, Pitman and Morgan F test, variance ratio
approximate, 4:263; 6:375; 7:20; 10:113, 410 table, 7:23
errors correlated, 1:71; 2:55; 6:265
F less than unity, 1:70; 2:56
F significantly small, 2:56
generalized, 7:61, and see multivariate analysis
interpretation of, see tests
non-normal data, 6:191
power of, 3:14; 7:6, 21, 125; 9:75; 10:267, 303
Fundamental curve, 4:197
Fungicides, 1:41, and see plant pathology
Future action, 1:8; 2:111; 5:297; 6:194, and see conclusions, inference, tests
Galton-Macalister distribution, see logarithmic normal distribution
Galton's quincunx, 2:97
Game theory, 6:190
Gamma distribution, 5:162, 301; 6:137, 194; 7:21, 360; 9:185, 375; 10:421, and see chi square
incomplete, 10:182
Gaussian distribution, see normal distribution
Gauss multipliers, 10:306
Generalized distribution, see multivariate distribution
Generating function, see characteristic function, cumulants, moments
Genetic advance, 2:77; 3:180; 6:204, 353; 9:90
Genetic combining ability, 2:78
Genetic correlation, 9:251, 339; 10:355
Genetic model, see models
Genetics, 1:7, 15, 55, 86; 2:28, 75; 3:55, 99, 123, 180; 4:1, 98, 138, 218, 254; 6:169, 175, 182, 186, 200, 207, 221, 344, 353; 7:18, 118, 270; 8:112, 173, 269, 388; 9:117, 205, 267; 10:164, 183, 184, 235, 405, 503, and see animal breeding, blood groups, cancer, cell division, chromosomes, crossing over, dominance, evolution, Drosophila, heritability, heterosis, inbreeding, linkage, mutation, natural selection, overdominance, path coefficients, plant breeding, random mating, recombination, selection-diffusion transformation, sex ratio, tetrads, viability
human, 3:193; 4:138; 8:112; 9:267; 10:184, 195, 309, 406, 445, and see birth rank
number of genes, 1:39; 2:78; 7:133
statistical, 1:27, 39; 9:125, 275, 336; 10:163, 164, 184, 307, 347

Genetic selection, 2:77; 3:180; 4:223; 6:104, 203, 208, 227, 353; 8:89, 176, 185; 9:90
tables for estimation of, 6:211
Genetic variance, 2:9; 4:254; 6:104, 141, 204; 7:75, 301; 8:89, 389; 9:337; 10:165, 312
Geometric series, 8:178
Gompertz curve, 7:247
Goodness of fit, 2:37; 5:320; 6:17, 146, 218, 361; 7:139, 396; 8:3, 18, 322; 9:122, 183, 279, 294, 322, 354; 10:33, 378, 396, 417, 460, and see chi square
power of chi square test of, 2:96; 8:272
Grading, see judging
Graeco-latin square, 8:171, and see orthogonal squares
Gram-Charlier series, 6:174
Graphics, 1:31, 41, 57, 73; 2:74, 83; 3:62, 151, 166; 4:199; 5:50, 83, 237, 339; 6:15, 120, 175, 180, 215, 328, 343, 345, 385, 432, 437; 7:200, 234, 260, 302, 328; 8:19, 120, 179, 191; 9:86, 201, 320, 463, 468, 476; 10:54, 150, 182, 187, 314, 538, 550, and see computation, nomography, probability paper, slide rule
Greenwood-Yule distribution, 9:375
Grouping, 1:38, 74, 76; 6:18, 171; 7:263; 8:19, 347; 10:538, and see chi square, Sheppard's correction
logarithmic, 8:22
Growth, 1:44, 73; 2:37; 3:58, 61, 172; 4:48, 139, 213, 240; 5:19, 27; 6:303, 310, 362, 395; 7:126, 135, 172, 247, 332; 8:271; 9:212, 260, 286; 10:302, 361, and see allometry, biometry, cell division, logistic curve, Robertson's equation, wound healing
compensation, 1:75
plant, 1:77; 2:37; 4:171; 10:282
G statistics, 2:96; 3:26, and see moments
Guard rows, see borders
Half-leaf method, 2:83
Half-plaid squares, 9:157, and see orthogonal squares
Harmonic analysis, 3:3
Hematology, 1:20, 37; 3:50, 172; 4:141, 143, 197, 200; 5:318; 6:184, 260, 282, 390, 435; 7:171, 185, 227, 331; 8:105, 390, and see blood counts, blood groups, erythrocytes, serology
Heredity, see genetics
Heritability, 4:223, 263; 6:104; 9:336; 10:312
estimation of, 9:90; 10:166
Hermite polynomials, 8:119
Heteroscedasticity, see scedasticity
Heterosis, 1:80; 2:75, 78; 3:180, and see overdominance
Histogram, 1:73; 5:220; 8:85; 9:124, and see class mark
bivariate, 2:97
Homoscedasticity, see scedasticity
Horticulture, 2:61, 76, 77, 78; 3:61, 139; 6:145, 164, 168; 7:300; 8:89; 9:39, 412, 428; 10:176, 387, 503, and see agronomy, perennials, tree crops
Hotelling's T, 7:61, 9:20, 259; 10:518, and see multivariate tests
Hypergeometric distribution, 7:276; 8:287, 391; 10:167, 308
Hypothesis, 1:38; 2:76, 82; 3:56, 61, 173, 181; 4:1, 80, 91, 249, 269; 5:162, 172, 274, 280; 6:122, 169, 259, 400; 7:114, 125, 354, 412; 8:35, 95, 96, 238, 276; 9:176, 244, 389; 10:33, 213, 311, 376, 517, and see models, tests
linear, 2:7, 22; 3:23, 95; 4:177; 5:73; 6:8; 7:1, 17, 71, 125; 8:2, 124; 9:11, 131, 229, 529; 10:199, 291
for variances, 7:17
non-null, 6:191
null, 1:8, 26, 70, 81; 2:13, 16, 18, 38, 68; 3:96, 98; 5:77, 81, 265, 339; 6:14, 178, 191, 299, 325, 372; 7:7, 17, 250, 299; 8:85, 112; 9:20, 28, 108, 416; 10:63, 92, 167, 170, 291, 408, 417, 448, 483, 487, 521
Identification, 5:88, 9:48
Immunology, 1:21; 3:172; 5:277; 6:320; 7:312; 8:177; 9:259; 10:171, 172, 173, 195, and see serology, tuberculin
Inbreeding, 1:27; 2:21, 75; 8:176; 9:344; 10:164, 360
Incomplete blocks, 1:12, 42; 2:30, 36, 42, 55, 78, 83, 115; 3:69, 176; 4:54; 5:1, 145, 285; 6:1, 44, 301; 7:117,

124, 125; 8:94, 95; 9:1, 31, 43, 59, 306; 10:375, 391, 487, and see lattices, randomized blocks, Youden square
doubly balanced, 7:117; 10:61
table, 10:82
balanced, 3:176; 4:103, 191, 213; 6:302, 362; 7:6; 8:66; 9:43, 242, 263; 10:61, 251, 301, 513
partially balanced, 3:178; 6:193, 302; 7:145; 8:66, 93, 122; 9:141, 261, 263; 10:273, 304, 511
table, 8:72
Incomplete experiments, 2:61, and see missing values
Incubation time, 6:344
Independent variable, see regression
Index of dispersion, 6:17, and see chi square
Induction, see inference
Industrial research, 5:175, 190; 6:228, 233, 290; 7:15, 102, 340; 8:17, 30, 75; 9:151, 259; 10:186, 187, 262, 273, 306, and see accidents, chemistry, physical science, statistical control
Infection interval, 5:243
Inference, 1:8, 70; 2:17, 104; 3:8; 4:218; 5:291; 6:189; 9:279; 10:308, and see conclusions, future action, tests
fiducial, 9:470; 10:100
Information, 1:38, 42, 78; 2:29, 43, 54, 122; 3:36, 42; 4:54, 185, 238; 5:69, 130; 6:2, 192, 194, 202, 221; 7:169, 219, 264; 8:11, 20, 33, 87, 170, 221, 334; 9:110, 175, 182, 198, 332, 469, 487, 523; 10:64, 132, 208, 286, 335, 405, 469, 539, and see analysis of variance, design of experiments, estimation, maximum likelihood, precision
expected, 7:269
interblock, 2:31, 35, 42, 55, 58, 115; 3:69; 4:62; 5:155; 6:26; 7:148, 287; 8:73, 93, 94, 122; 9:59, 146
intrablock, 2:35; 3:69; 4:62, 109; 6:26, 45; 7:148; 8:77, 122
matrix, 6:188, 213; 7:250, 268; 9:261
observed, 7:269
theory, 9:523
Instrumental variate, 7:34; 8:269
Integral equations, 3:58; 4:139; 9:461
Integration, 8:21
Interaction, 1:11, 15, 16, 31, 37, 41, 55, 85; 2:24, 26, 30, 62, 77; 3:10, 40, 50; 4:54, 101, 132, 212, 243; 5:72, 145; 6:15, 28, 51, 79, 234, 318, 366; 7:13, 18, 324; 8:5, 58, 264; 9:3, 171, 253, 264, 299, 414, 434; 10:21, 102, 165, 187, 255, 267, 494, 505, and see analysis of variance, main effect, models
as error term, 2:43; 3:91; 4:133; 6:10, 319; 7:14, 109; 8:75, 238; 9:256, 337, 416; 10:331, 346
generalized, 4:57
interpretation of, 9:388; 10:21
minimized, 3:51
of quality and quantity, 7:433; 9:387
Intercept distribution, 4:9
Interpolation, 2:121; 4:197; 8:51; 10:4, and see extrapolation, moving averages, theory of errors
Invariance, see information, variance
Irrigation, 9:157
Iteration, 1:39; 2:121; 3:166; 6:10, 182, 223, 305; 9:179, 278; 10:156, 320, and see computation, least squares, maximum likelihood
J-shaped distribution, 3:60; 7:376
Judging, 1:43, 48; 2:17, 76; 5:77; 6:1, 206; 9:107; 10:302, 305, 375, 524, and see organolepsis
contests, 2:67
number of samples, 6:1; 9:31
Kapteyn's distributions, 3:174
Karber's method, 4:197; 5:85; 6:321, 413, 432; 7:295; 10:150
Kurtosis, 6:103; 8:332; 9:11; 10:429, and see moments
Lagrange multipliers, 5:194; 7:28; 10:493
Largest value, see extreme deviate
Latent roots, 7:61; 9:389, and see matrices
Latin rectangle, 3:70, and see Latin square
analysis of variance of, 3:72
Latin square, 1:13, 26, 31, 42, 48, 77; 2:35, 41, 83, 87; 3:25, 28, 32, 70, 139; 4:6, 61, 118, 123; 5:1, 71, 102, 179, 305; 6:202, 289, 301, 318; 7:6,

261, 319; 8:169, 238; 9:3, 31, 158, 416; 10:155, 184, 239, 262, 290, 400, 507, and see cross-over design, Latin rectangle, orthogonal squares, quasilatin square
incomplete, 3:70; 6:302, and see Latin rectangle, in time, 2:87
effect of previous treatment, 2:89; 8:170
systematic, 7:167
Lattices, 1:13, 42; 2:36, 42, 55, 78; 3:176; 4:54, 109; 5:144; 6:1, 34, 301
balanced, 1:13; 4:108; 6:301; 9:43
cubic, 1:13; 4:112; 5:144; 7:146; 8:123
incomplete square, 6:34
latinized rectangular, 8:73; 9:32, 43
near-balance rectangular, 9:101
quadruple, 6:301
rectangular, 2:35, 115; 5:85; 6:25, 301; 7:145; 8:93, 94, 123; 9:43, 101; 10:292, and see latinized rectangular, near-balance rectangular, triple rectangular
unequal blocks, 7:287
simple, 1:12, 13; 6:301; 7:145; 8:93; 9:59
square, 1:12, 13; 2:42, 58, 78; 4:115; 6:301; 7:146; 8:94, 123
three-dimensional, see cubic
triple, 1:13; 2:36, 58; 4:60, 110; 6:301; 8:135
triple rectangular, 4:137; 5:1, 85; 7:145; 9:422
LD 50, see ED 50
Learning theory, 4:98; 9:280
Least significant difference, 1:80; 2:31; 3:79, 92; 4:214; 6:14; 9:43, 147; 10:167
Least squares, 1:36, 57, 78; 2:22, 37, 43, 94; 3:3, 8, 11, 163; 4:27, 178, 240, 248; 5:207, 228, 246; 6:10, 120, 177, 180, 187, 315, 397, 414; 7:106, 118, 125, 233, 247, 309; 8:43, 91, 124, 249, 384; 9:122, 131, 235, 291, 499; 10:29, 64, 108, 181, 227, 255, 292, 304, 335, 400, and see adjusted means, analysis of variance, curve fitting, estimation, fitting constants, fitting regression line, matrices, normal equations, residuals, theory of errors
damped, 9:505
iterative computation, 9:205, 291
Lee's phenomenon, 1:76
Life tables, 3:58, 59, 129; 4:51, and see acturial statistics, demography, Gompertz curve, mortality
Life, expectation of, 8:91
Likelihood, 7:28; 10:100, and see maximum likelihood
Likelihood-ratio estimate, 5:80; 6:378; 7:19, 125; 8:95, 185; 9:20; 10:63, 377, 492
Limnology, 5:14, 28, and see fish, plankton
Linear dependence, 2:23; 7:269, and see matrices
Linearity, 1:9, 10, 38; 3:149; 5:88, 261, 282, and see regression, tests
test of, 1:75; 3:137; 4:104; 5:88, 211, 266; 7:210; 10:213, 423
Linkage, 1:28, 85; 4:265; 5:305; 7:270; 8:173; 10:184, and see crossing over
Linked block design, see chain blocks
Logarithmic distribution, 5:162, 165; 7:121; 8:178; 9:177
Logarithmic normal distribution, 4:22; 6:103, 104, 147; 8:271, 273, and see normal distribution
discrete, 6:103
Logistic curve, 1:39, 44; 3:173; 4:200; 6:414; 7:247, 327; 8:249, 268; 9:266, 271; 10:178, 227, 310, 398, and see growth, logit transformation, populations
Longevity, see life tables
Long-term experiments, 1:16; 2:61; 6:194, 241, 315; 7:21; 8:14; 9:428, 429; 10:324, and see field experiments, perennials, rotation experiments, tree crops
Mail surveys, 9:265, and see sampling
Main effect, 1:31; 5:145; 6:315; 7:14; 8:5, 264; 9:389, and see analysis of variance, interaction
Markov, 6:339; 9:276; 10:361, and see stochastic processes
Masking, 9:17
Mass action, 7:332, 435
Matching, see balancing, paired comparisons

Mathematical biology, see biometry
Matrices, 6:251, 335, 372; 7:269, 284; 8:6; 9:49, 276, 499, 533; 10:365, 453, 518, and see covariance matrix, Doolittle method, error matrix, Gauss multipliers, information matrix, Lagrange multipliers, latent roots, linear dependence, multivariate analysis, simultaneous equations
  inverse, 4:178, 245, 248; 6:180, 380; 8:5, 339; 9:210, 237; 10:10, 306, 346, 370, 519, 534, 547
  square-root method, 8:9
  stochastic, 10:347
Maximum likelihood, 1:28, 76, 78; 3:26, 97; 5:83, 85, 86, 130, 166, 207, 305; 6:19, 104, 110, 129, 171, 187, 188, 194, 212, 224, 361, 413, 432; 7:4, 29, 120, 136, 240, 250, 268, 295, 328, 404, 8:1, 16, 95, 179, 249, 271, 279, 292, 316, 390; 9:33, 59, 96, 177, 194, 198, 261, 279, 291, 321, 365, 467, 485, 494, 501; 10:131, 178, 184, 227, 303, 377, 395, 402, 460, 488, 532, 545, 557, and see efficiency, estimation, information, minimum unlikelihood, theory of errors
  iterative computation, 5:83, 85, 131; 6:104, 171, 212, 414; 7:29, 136, 268; 8:24, 95, 179, 281, 317; 9:60, 96, 467, 476, 489, 503; 10:145, 403, 491, 533
Mean, 2:73; 3:26, 43, 163; 5:288; 6:60, 261; 7:9, 181; 9:75, 220, 523; 10:198, 307, and see expectation, median, mode, moments
  adjusted, 1:11; 2:15, 31, 35, 36, 53, 78, 118; 3:34, 69, 77, 84; 4:159, 177, 244; 5:5, 156; 6:27, 46; 7:313; 8:75; 9:146, 10:70, and see least squares
  geometric, 6:146
  harmonic, 5:306; 8:304
  of extremes, 5:288
  of function, 7:68
  semiweighted, 10:109
  weighted, 2:55; 3:78; 8:304; 9:59, 388, 447; 10:108, 556, and see semiweighted table, 10:110
Median, 2:74; 6:190; 8:30, 201; 9:75; 10:539, and see mean, mode
Median effective dose, see ED 50
Medicine, 1:3, 52; 2:47, 98; 3:99, 109, 123; 4:137; 5:81, 95, 127, 243, 277; 6:181, 259, 303, 322, 395, 435; 7:126, 309, 320, 340, 435; 8:90, 105, 113, 277; 9:260, 320, 431; 10:160, 163, 171, 195, 313, and see accidents, age distribution, anthropology, cancer, case histories, chemotherapy, clinical, demography, diagnosis, digitalis, epidemiology, genetics, hematology, immunology, morbidity, mortality, nervous system, pharmacology, physiology, psychology, public health, serology, toxicology, tuberculosis, virology, wound healing
Metameter, 5:261, 282; 6:101; 7:212, 327; 8:18, 324; 10:305, 306, 531, 544, and see metric
Metric, 1:27; 3:173; 4:4; 6:145; 9:11; 10:130, and see metameter, scales
Microbiological assay, 3:106; 5:280; 8:376, and see bacteriology, bioassay
Miller disc, 6:390
Minimax estimation, 6:188, 194; 10:36, 100
Minimum chi square estimate, 1:39; 5:85, 86, 166; 6:414, 432; 7:338; 9:376; 10:148, 178, 227
  computation of, 5:86
Minimum unlikelihood, 10:100
Minimum variance estimate, 3:163; 4:66; 6:28, 110, 224; 7:157; 8:92, 269; 9:91; 10:400
Missing values, 1:13, 14, 37; 2:35, 41, 94, 116; 3:94; 4:125, 241; 6:188; 8:42, 383; 9:256, 263, 425; 10:273, 298, 400, and see censored data, dummy variates, fitting constants, rejection of data
  computation of, 2:41; 3:28; 8:46
Mitscherlich's law, 7:248; 9:498
Mode, 1:73; 6:103; 9:191, and see mean, median
Models, see biometry, components of variance, hypothesis, learning theory, peck order, regression, structure, theory of errors, transformations, variance related to mean
  biological, 9:176, 275

genetic, 4:255; 8:389; 9:96, 340; 10:165, 237, 307
least squares, 10:304
mathematical, 1:27, 38; 2:74; 3: 4, 19, 104, 137, 145; 4:49, 135, 140, 201, 225; 5:65, 191, 214, 244, 261, 319; 6:169, 181, 186, 250, 330, 342, 353, 367, 437; 7:17, 99, 105, 115, 134, 185, 227, 269, 300, 335, 408; 8:2, 14, 173, 238, 264, 268, 271, 286, 315, 390; 9:32, 59, 75, 92, 131, 203, 212, 290, 305, 388, 412, 427, 465, 471, 529; 10:62, 102, 305, 310, 375, 390, 400, 408, 459, 480, 487, 517, 541
physical, see mathematical
probability, 6:286; 8:267
statistical, 3:162; 4:168; 5:76; 6:434; 7:134; 8:217; 10:236
stochastic, 9:278: 10:361
Moments, 2:13, 68, 96; 6:28, 374; 7:140, 277, 360; 8:18, 278, 326; 9:179, 360; 10:142, 152, 199, 397, 429, 464, 543
factorial, 3:58; 7:360, 402
generating function, 2:13; 6:30; 8:302, 328; 9:223; 10:464, 543
factorial, 7:121, 391
method of, 6:122; 7:142, 404; 8:20, 280, 292, 329; 9:179, 376, 485, 494; 10:548
Morbidity, 1:3; 3:58, 109; 4:51; 5:8? 243; 6:170, 172; 8:174
Mortality, 1:4, 12, 76; 3:58, 109, 175; 5:24, 34; 6:170, 252; 7:120, 155; 8:91, 173, 271, 272, 300, 308, 339; 9:212; 10:308, 313, and see acturial statistics, census, life tables
Most probable number, 6:105, 121, 127
Moving averages, 6:318, and see interpolation, time series, trend
estimation, 4:197; 6:414, 432; 7:302; 8:51, 250
Multifactor design, 9:304
Multinomial distribution, 2:12; 6:20, 186; 7:115, 269; 8:62, 114, 295, 389; 10:184, and see binomial
Multiple classification, 2:21; 7:1
Multivariate analysis, 3:3, 51; 4:50; 5:78, 81; 6:190, 259, 303, 372; 7:37, 106; 9:20, 259; 10:208, 518, and see analysis of variance, discriminant function, D statistic, factor analysis, Hotelling's T, matrices, multivariate distribution, multivariate tests, quadratic forms
computation, 6:268, 305, 373; 7:39; 10:209, 519
confidence limits, 7:53
homogeneity of error matrices, 6:386
Mutation, 7:133, 336; 8:86; 9:260; 10:373, and see radiology
social, 6:337
Natural selection, 1:7, 74; 6:353; 7:119, and see evolution, genetics
Nervous system, 3:182; 4:91, 140, 144, 153; 7:123; 8:37, 174, 267, 389; 9:279, 305, and see neural net, physiology, psychology, reaction time
Neural net, 4:93; 6:183; 9:279, and see model, nervous system, random net
Newton's method, 9:321
Nomography, 1:32, 40, 43; 3:62; 5:322, 330; 7:200; 10:320, and see computation, graphics, slide rule
table for construction, 5:334
Non-numerical data, 2:123; 3:50; 5:79, and see coding, contingency, scores
Non-orthogonal design, 10:513
Nonparametric tests, 1:80; 3:119; 6:177, 407; 7:36, 116; 8:14, 29, 33, 295; 9:109, 260; 10:62, 303, 532, and see order statistics, ranks, sign test
efficiency of, 1:81; 6:282
of trend, 6:282
Normal distribution, 1:74, 124; 3:14, 22; 4:95; 5:325; 6:145, 174; 7:233, 337; 8:18; 10:227, 429, and see Galton's quincunx, Hermite polynomials, logarithmic normal distribution, normality
bivariate, 1:71, 86; 2:101; 4:21, 50; 6:410; 9:261, 460; 10:542, 550
circular, 9:460
cumulative, 1:44; 8:268
multivariate, 10:518
truncated, 5:84; 6:187; 8:19, 319; 9:122, 261, 489
tables, 9:490; 10:539
Normal equations, 2:22, 35; 3:177; 4:178, 240; 7:11, 236, 249, 338; 9:293, 393; 10:65, 268, 274, 303, 336, and

see least squares, matrices, observation equations
Normality, 5:261, 279, and see g statistics, normal distribution
departure from, 3:26; 5:265, 279
deviation from, 8:117
non-normality, 3:23, 32; 5:109; 6:145, 191, 408; 7:6
test of, 1:76; 2:96; 8:18; 10:544
Nuisance parameters, 7:5, 19, 125; 10:304, 378
Nutrition, 2:27, 58; 4:49, 213, 267; 5:178, 250, 289; 6:282, 304, 319; 7:39; 8:92, 387; 10:302, and see vitamins
Obituaries, 2:31, 4:276; 7:221; 9:114
Observational equations, 2:7, 22; 9:392, and see least squares, normal equations
Odor, see organolepsis
Operating characteristic, 5:84; 6:63; 7:365; 8:300, and see sampling, statistical control, tests
Opinion research, 9:286, and see psychology, rumors, sampling
Optimum catch, 5:14, 213, and see fish
Optional stopping, 10:94
Order statistics, 6:177, 192; 8:40, and see extreme deviate, nonparametric tests, ranks
Organic variation, see biological variation
Organolepsis, 2:123; 4:80; 5:337; 6:1, 102, 206; 7:64; 8:75, 95, 96, 218; 9:22, 39, 107, 264, 296; 10:61, 167, 375, and see fatigue, food technology, judging, masking, psychology, scores, taste blindness, triangle test
flavor, 2:13; 3:104; 6:1; 9:1; 10:391, 521
number of samples, 6:1
odor, 2:11; 6:1
Ornithology, 3:59
Orthogonal arrays, 6:179, and see orthogonal squares
Orthogonal functions, 2:28, 29, 61; 9:395; 10:338
Orthogonality, 7:33, 47; 9:227, 260, 304, 468, 531; 10:184, 240, 255, 328, 341, and see comparisons
non-orthogonality, 1:77; 5:72; 9:227; 10:513
Orthogonal polynomials, 1:10, 20; 2:29, 61, 132; 6:180, 317, 384; 9:306, 392, 429, 434, 473; 10:171, and see curve fitting
Orthogonal squares, 6:180, and see Graeco-latin square, half-plaid square, Latin square, orthogonal arrays
completely, 4:111; 7:168; 8:73; 9:102
Overdominance, 4:263; 6:104, 219; 7:18
Pairing, see balancing, paired comparisons
Palatability, see organolepsis
Parallelism, test of, 1:58; 3:80; 4:104; 5:266; 7:210, and see regression
Parallelogram design, 9:467
Parasitology, 3:89; 6:59, 345; 7:310; 9:190, 460; 10:107
Partitions, 1:81; 6:19; 10:362
tables, 1:81
Path coefficients, 1:75; 9:347; 10:347, 359
Pearsonian distributions, 3:25, and see gamma distribution, Type VI distribution
Peck order, 6:104, 330; 9:283
Percentages, see proportions
Percantile, 2:74, and see quantile
Perennials, 2:61; 3:139; 10:503, and see field experiments, long-term experiments, tree crops
Permutation tests, 10:303, and see nonparametric tests
Pharmacology, 1:57; 2:34, 58, 82, 87, 95; 3:173; 4:153, 269, 271; 5:83, 289; 6:322; 7:123, 171, 227; 8:206, 218, 232; 9:263, 305, 431; 10:160, 246, 435, and see analgesia, chemotherapy, digitalis, selection of dosages, toxicology
Phylogeny, 7:121, and see taxonomy
Physical science, 2:103; 7:102, 257, 330; 8:386; 9:127, 275; 10:19, 103, 273, and see Brownian motion, chemistry, industrial research, radiology, spectrography
Physiology, 2:55; 3:65, 99, 173; 4:49, 100, 141, 153, 163, 269; 5:49, 71; 6:169, 172, 183, 395; 7:122, 185, 227, 275, 284,

301, 336; 8:37, 175, 186, 206, 267, 389, 390; 9:267, 431; 10:221, 307, 313, 442, and see endocrinology, medicine, nervous system, nutrition, respiration, threshhold
phytopathology, see plant pathology
Pitman and Morgan F test, 1:71; 2:55
power of, 1:71
Placebo, 2:87, 95; 8:218, and see selection of experimental units
Plankton, 1:49; 6:142; 9:190
Plant breeding, 1:13; 2:28, 36, 75, 77; 3:180; 4:75, 120, 261; 5:252; 6:146, 166, 200, 226, 318; 7:18, 39, 75; 9:39; 10:240, 298, and see animal breeding
Plant pathology, 1:41; 2:81; 6:344, and see fungicides
Plant physiology, 6:174
Plot size and shape, 6:202; 7:300; 9:158, 414; 10:175, 503, and see block size, borders
Poisson distribution, 1:43, 44, 51, 76, 85; 2:2, 76, 82, 97; 3:32, 41, 90, 166; 4:95; 5:83, 129, 162, 165; 6:17, 180, 193, 291, 391; 7:355, 395; 8:55, 178, 287, 384; 9:176, 209, 320, 354, 427; 10:135, 167, 224, 229, 248, 304, 419, 486, 524
compound, 7:400; 8:275
double, 9:192
modified, 6:173
Thomas double, 6:103; 9:192
truncated, 4:137; 8:275; 9:485; 10:402
tables, 10:405
Polya distribution, 2:76; 6:103; 9:192
Pooling, 1:61, 63; 2:58; 3:25, 31; 4:272; 5:176; 6:270, 285; 7:79, 289; 9:339; 10:106, 458
multivariate analysis, 6:386
Populations, 1:7, 11, 27, 44, 73, 77; 2:1, 48, 76, 3:58, 59, 113, 145, 193; 4:21; 5:76, 78, 80, 115, 213; 6:103, 146, 219, 250; 7:275, 300, 301, 332; 8:17, 269, 271; 9:47, 109, 117, 212, 278, 336; 10:362, and see demography, distributions, fish, statistical genetics
estimates, 3:145; 8:89, 272, 391; 10:181
finite, 1:69; 2:11; 3:119; 5:205; 7:3, 98, 156, and see sampling
management, 1:31, 75; 5:14, 116, 213, and see ecology
virtual, 5:32
Poultry, 1:11, 25, 62; 2:34; 3:89; 4:213, 240; 5:176, 178, 250, 264, 277; 7:126, 215; 9:256; 10:103, and see peck order
egg production, 2:56; 4:223
egg shape, 9:336
Precision, 1:10, 14, 30; 2:28, 53, 78; 4:108, 127, 197, 230; 5:68, 142, 189, 307, 331; 6:107; 7:85, 102, 203, 315; 8:85, 170, 177, 244; 9:59, 158, 177, 293, 452, 531; 10:109, 261, 286, 294, 301, 503, and see information
of measurement, 7:36; 8:235
Prediction, 1:87; 2:61, 102; 3:181; 4:13, 50, 141, 184, 247; 5:14, 38, 207; 6:398; 7:75, 365, 374; 8:29, 272; 9:47, 261, 530; 10:309, and see conclusions, inference, judging, regression
accuracy of, 6:184
Preferences, see organolepsis
Probability, 1:81; 2:12, 38, 49, 82, 106; 3:119, 162: 4:2, 99; 6:109, 163; 6:20, 63, 106, 128, 332; 7:134, 180, 276, 392; 8:91, 173, 199, 222, 279; 9:54, 177, 266, 276, 321, 523; 10:63, 94, 131, 229, 402, 459, and see characteristic functions, distributions, proportions, random variables, sampling, tests
combination of, 9:16
conditional, 8:295; 10:347, and see conditional distribution
generating function, 7:391; 9:277, 357
inverse, 5:292; 6:109
paper, 2:74; 5:339; 6:180; 7:302, 328
posterior, 10:100
transition, 10:348, 361
Proportions, 1:42, 43; 3:45, 56; 4:211; 6:178, 298, 390; 8:1, and see binomial distribution
Pseudo-effects, 5:145
Pseudofactor, 9:297
Pseudofactorial design, 2:115; 4:54, 110; 5:1, 145; 6:35; 9:299
Pseudovariate, see dummy variates

Psychology, 2:73, 87, 103, 124; 3:99; 4:98, 140; 5:76, 78, 79, 80; 6:1, 184, 262, 435; 7:36, 114, 275, 341; 8:169, 218, 268, 290; 9:4, 287; 10:94, 302, 313, 442, and see accidents, attenuation, case histories, communications, factor analysis, judging, learning theory, medicine, nervous system, opinion research, organolepsis, peck order, placebo, reaction time, reliability, rumors
Public health, 1:52; 3:110, 123; 5:82, 95, 96; 7:309; 8:174; 10:119, 185, and see epidemiology, medicine
Quadratic forms, 6:185, and see multivariate analysis
Quadrat sampling, 1:31, 47; 5:163; 9:189, 369; 10:402, and see ecology
Qualitative data, see non-numerical data
Quality, see interaction of quality and quantity, organolepsis, statistical control
Quantal response, 1:60; 2:34, 95; 3:62, 175; 4:91, 197; 5:83, 274, 288; 6:328; 7:120, 333; 8:19, 95, 179; 9:467; 10:130, 178, 185, 306, 534, 556, and see bioassay, models
Quantification, see scales
Quantile, 8:30; 9:47, and see percentile, quartile
Quantitative analysis, 3:61; 4:234; 5:195, 329; 6:118; 7:102; 8:85, 264; 9:87, 135; 10:407
Quartile, 3:56; 8:288, and see quantile
  computation of, 3:57
Quasifactorial design, see factorial experiments, lattices
Quasilatin square, 9:158
Radiology, 2:97; 3:60; 4:139; 6:123, 169, 184; 7:126, 336; 8:86, 90. 339; 9:201, 280; 10:282, 371, 455, and see mutation, Poisson distribution
Randomization, 2:18, 29; 3:19, 24, 33, 72, 176; 5:145, 178, 279; 6:35, 201, 317; 7:47, 288, 309; 8:75, 171, 189, 219, 234; 9:300; 10:184, 299, 522, and see selection of data
  angular, 9:260, 305
  restricted, 6:323
Randomized blocks, 1:13, 16, 31, 42, 85; 2:7, 15, 16, 31, 36, 41, 54, 61, 78; 3:25, 40; 4:54, 103, 192, 234, 244; 5:145, 169, 174; 6:44, 301; 7:115, 316; 8:42; 9:31, 306, 416; 10:400, 516, and see balancing, incomplete blocks, interblock error, interblock information, split plots, systematic designs
Random mating, 4:254; 7:18
Random net, 9:280
Random numbers, 5:106, and see random sampling
Random process, 4:3; 8:173, and see stochastic processes
Random variable, 2:106; 3:9, 57, 165; 5:191; 6:14, 190, 193, 194, 284; 7:18; 8:13, 124, 287; 9:49, 275, 469; 10:178, 460
Random walk, 1:7; 9:281, 460, and see random net
Range estimate of variance, 2:73; 5:288, 331; 6:437; 7:222; 9:79, 523
  table, 9:89
Range, significance of, see tests
Ranks, 1:44, 80; 2:38, 67, 76, 3:47, 104, 119, 138; 6:1, 177, 282; 7:114, 125; 8:33, 95, 96; 9:31, 45, 108, 260, 264; 10:63, 375, 391, 487, and see birth rank, nonparametric test, order statistics, scores, transformations
  rank tests, power of, 9:32
  tables, 1:81; 3:121; 6:285; 8:37
  tied, 1:80; 8:38, 114; 9:109; 10:64
Rate, average, 3:55
Reaction time, 3:172, 182, and see nervous system, psychology
Recombination, 1:28; 4:1; 6:221; 10:164, and see linkage
Rectangular distribution, 3:119, 194; 8:116; 10:182
Recurrence formulas, 9:356
Reed-Muench method, 4:197; 6:413, 432; 7:295; 10:147
Regression, 1:15, 27, 31, 58, 75, 86; 2:15, 27, 37, 83, 101; 3:3, 8, 36, 47, 55, 57, 97; 4:22, 157, 182, 247, 255; 5:68 87, 123, 140, 175, 237, 245, 261; 6:3, 184, 329; 7:64; 8:19, 115, 384; 9:90, 392; 10:212, and see adjusted means, canonical analysis, correla-

tion, covariance, error regression, fitting regression line, orthogonal polynomials, prediction, scatter diagram, scedasticity, statistical control, structure, trend
asymptotic, 5:320; 7:247; 9:498, and see exponential regression
tables, 7:254; 9:514
adjustment, 1:30; 2:53; 3:34, 103; 7:186; 9:267, 312;
analysis, 3-61; 6:294; 7:33; 8:89, 115; 9:90, 170, 267, 304; 10:56, 173, and see analysis of covariance
both variables subject to error, see fitting regression line
coefficient, 2:102, 105; 5:69; 6:168; 7:45; 8:269; 10:178
biased, 9:528
nonsignificant, 1:59
coefficients equal, 4:250
curvilinear, 1:9, 10, 31, 41, 44; 2:25, 28, 61, 63; 3:137; 4:106, 132, 175, 182; 5:51, 68, 282; 6:5, 120, 166, 294, 384; 7:234, 301, 324; 8:89; 9:307, 388, 401, 445, 473; 10:184, 558, and see asymptotic, exponential hyperbolic
dependent variable, 2:15, 4:27; 5:73; 6:303; 9:267
deviations from, 1:86; 3:8; 5:179; 6:296
direct, 2:109
equation, 3:103; 4:247; 5:228; 6:5; 7:187, and see regression line
exponential, 3:149; 7:126, 135, and see asymptotic regression
heterogeneity of, 1:11; 2:28; 3:61; 4:132, 167; 6:329; 7:193; 9:475
homogeneity of, 3:66
hyperbolic, 2:63
independent variable, 2:15, 109; 4:27, 49; 5:73, 179
affected by treatments, 6:166, 303; 7:51
integral, 2:29
intercept, 3:61, 66; 5:229; 6:210; 7:261; 8:89
internal, 7:187
inverted, 2:109
line, 2:105; 4:159; 8:269; 9:119, and see regression equation
model, 8:374
multiple, 3:3; 4:50, 160; 5:119; 6:185, 251; 7:33, 64, 118, 188, 249, 301; 8:95, 372; 9:467, 528; 10:29, 306
nonlinear, see curvilinear
of variance, 9:267
on time, 2:29, 61; 6:282
partial, 2:23, 29; 5:175; 6:168, 293; 7:198; 9:528; 10:213
polynomial, 2:28; 3:137; 6:180, 230, 317, 437; 7:188, 247; 9:306; 10:4, 28
selected data, 2:103; 3:55
transformation, 1:43
weighted, 10:434
Rejection area, 1:70; 2:17; 6:63; 8:185, 299; 9:25, 10:311, and see acceptance area computation of, 6:62; 8:185
Rejection of data, 2:41, 95; 3:27; 4:270; 8:243; 9:12, 74, and see gross errors, missing values, selection of data
test for, 3:27; 9:84
Reliability, 7:61
Replication, 1:13, 50; 2:18; 3:61, 176; 4:54, 110, 215, 237; 5:178, 195; 6:290, 302; 9:148; 10:175, 506
fractional, 4:59; 5:73; 6:228, 233, 319; 9:295; 10:4, 45, and see aliases, defining contrast
hidden, 6:234
in time, 1:16; 6:229, and see cross-over design, Latin square in time, rotating factors
Reproductive rate, see birth rate
Research in statistical methods, 1:6, 23, 53; 2:33; 3:24; 6:75, and see sampling studies of statistical problems
Residual effects, see carry-over effect
Residuals, see theory of errors
Respiration, 6:183
Response rate, 9:265; 10:91
Response surface, 10:16
Response time, see time response
Reversal experiments, see cross-over design
Risk function, 10:169, 318
Robertson's equation, 2:37
Rotating factors, 6:312, and see replication in time
Rotation experiments, 10:324, and see long-term experiments

Rumors, 8:268; 9:285, and see epidemiology, opinion research, psychology
Sample size needed, 2:35, 95, 120; 6:107, 167, 174, 182, 294, 325, 402, 424; 7:7, 157; 10:32, 168, 317, 420, and see average sample number, organolepsis
tables, 10:320
Samples, small, 6:113, 414; 7:136; 10:485
Sampling, 1:43, 47, 49; 2:48, 77; 3:8, 19, 163, 193; 6:59; 7:83, 97, and see census, components of variance, design of experiments, distributions, field sampling, mail surveys, operating characteristic, opinion research, quadrat sampling, response rate, soil
balanced, 8:110
correlation in, 5:172
cost of, 1:34, 67; 2:2, 10; 4:231, 237; 5:189; 6:402; 7:93, 157
double, 1:29; 6:66; 7:275; 10:94, 167, 317
ecological, 3:158
error, see variance
experiments, see sampling studies
grab, 4:234
inverse, 6:291; 7:300; 8:286; 10:89
multiple, 8:286
multivariate distribution, 8:94
nested, 5:189
optimum, 4:239; 5:189; 7:86; 8:270
panel, 8:270
probability, 8:91
random, 1:29, 77; 3:101; 4:223; 8:94
schedule, 7:95
selective, 1:76; 2:48; 3:101, 193
sequential, 6:59; 7:300; 8:286; 9:191; 10:89
stratified, 1:48; 3:158; 5:174, 205; 7:155, 301; 10:103
studies of statistical problems, 3:24; 5:106; 7:338; 9:87, 447; 10:178, 303, 305, 466
surveys, 3:117; 5:82, 165; 8:10, 92, 100, 183, 245, 386; 10:91, 479
systematic, 1:29, 48; 2:3; 3:102; 4:223
time series, 9:264
Sanitary engineering, 6:117, 127, and see bacteriology, chemistry
Scales, 3:50; 5:77, 80; 6:1, 145; 7:114, 122; 9:29, 109; 10:166, 481, and see fish, metric, scores
Scatter diagram, 5:30, 41, 51, 122, 172, 230, 238, 247
Scedasticity, 3:12; 4:25; 5:261, 283; 6:10, 101, 377; 7:366; 9:471, and see homogeneity of variance
Scores, 1:43; 2:11, 68, 124; 4:272; 5:78, 340; 6:1, 178, 213, 221, 227; 7:114, 268, 380; 8:7, 112; 9:1, 24, 40, 118, 180, 197, 264, 267; 10:131, 184, 309, 376, 435, and see discriminant function, organolepsis, ranks, scales
analysis of, 3:50; 9:29; 10:63
for ranks, 1:80; 3:49; 9:29
Selection, see balancing, design of experiments, genetic selection, natural selection, optional stopping, randomization, rejection of data, sampling
of comparisons, 1:26; 3:28, 139, and see tests of results suggested by data
of data, 2:103; 3:55; 5:46; 7:47, 104; 8:243; 9:193, 265; 10:387, 549
of dosages, 5:83, 289; 8:219, 246; 9:263, 306, 332, 428, and see selection of levels
of experimental units, 1:31; 3:19, 175; 4:211; 5:178, 279, 301; 8:169, 214, 234, and see placebo
of judges, 4:81
of levels, 7:252; 9:428, and see selection of dosages
of method of statistical analysis, 6:319
of parameters, 8:17
of significance level, 1:50; 8:196
Semiweights, 10:109
Sensory tests, see organolepsis
Sequential experiments, 2:113; 6:59, 228, 230, 245; 8:183, 188, 286; 9:24, 262, 428, 523; 10:5, 30, 89, 169, 317, and see average sample number, design of experiments
Serology, 1:20, 37; 3:50, 70, 173; 4:141, 143, 197; 5:277, 289; 6:323, 345; 9:61; 10:172, 445, and see blood groups, complement fixation, hematology, immunology

Service, computing and consulting, 1:6, 23, 47, 53, 83; 2:33, 71; 5:256
  presentation of results, 10:54
Sex ratio, 2:58; 3:100; 5:49, 86; 8:388; 9:212; 10:93
Sheppard's correction, 1:76; 8:18; 10:157
Shrimp, 8:323
Significance, see selection of significance level
  nonsignificance, 1:8; 2:113; 6:143, 294
Significant cut, 7:50
Sign test, 1:81; 2:95; 8:295; 9:109; 10:487, and see nonparametric tests
Simultaneous equations, see matrices
  nonlinear, 1:11
  solution, 2:23; 4:240
Skewness, 2:68, 73; 3:25, 174; 5:263, 287; 6:14, 103, 114, 174; 8:332; 9:11; 10:183, 196, 429, 543, and see moments
Slide rule, probit, 1:46, and see computation, nomography
Sociology, 2:70; 5:77, 86; 9:283, and see communications, economics, mutation, opinion, peck order, psychology, rumor
Soil heterogeneity, 1:13; 3:34; 4:113; 6:318; 10:103, and see agronomy, field experiments
Soil sampling, 1:77; 5:174; 8:386, and see field sampling
Soil science, 1:27, 30; 2:5, 28, 64; 3:29; 7:301; 9:189
Soil treatment, 10:291
Spectrography, 9:135
Split plots, 1:14, 16, 31; 2:43; 4:118, 211; 6:164, 301, 368; 9:31, 157, 298, 429, 533
Staggered design, 4:195
Standard deviation, see variance
Standard error, see variance
Standard variety, 1:13, 14; 2:30; 6:202, and see check plot
Statistical control, 4:157; 5:77, 175; 6:435; 8:169; 9:260, and see analysis of covariance, concomitant variables, control chart, cross-over experiments, operating characteristic, sampling
  of gradient, 10:282
  of quality, 1:10; 4:125; 5:83; 6:229; 7:15; 9:259; 10:317
Statistics curricula, 1:5, 22, 33, 52, 65, 83; 2:32, 70, 98; 3:179; 4:87, 220; 5:94, 96, 344; 6:87, 231; 7:130, 449; 8:100; 9:116, 273, 518, 538; 10:160, 161, 162, 192
Steepest descent, method of, 9:65, 10:6, 30
Stochastic processes, 3:164, 174; 4:99; 5:89; 6:92, 219, 251; 7:121, 358; 9:212, 260, 275; 10:22, 201, 347, 361, and see correlated errors, economics, Markov, random processes, time series
Structure, 7:36; 10:308
Studentized extreme deviate, 5:107
Student's t, see t test
Subjective evaluation, see judging, organolepsis, psychology
Successive approximation, see iteration
Sufficient statistics, 6:17; 7:338; 8:173, 292, 391; 9:463; 10:150, 170, 178, and see efficiency, estimation
Surveys, see sampling
Survival curve, 9:201; 10:371, 531, and see wear curve
Survival time, 3:175; 6:99, 104, 120; 8:37, 90, 271, 308; 10:555
Switchback design, see cross-over design
Symmetry, 5:81; 6:259
Systematic designs, 6:318; 7:167, 309; 8:169, and see randomization
Tables, miscellaneous, 1:69; 2:4, 12, 68, 69; 3:41; 5:86, 106; 6:115, 354; 7:219, 281; 8:200, 279, 305; 9:324, 450, 487; 10:29
  graphical, 6:426; 7:206; 10:55, 320
Taste blindness, 1:21
Taste tests, see organolepsis
Taxonomy, 1:39, 55, 74; 6:184; 7:182; 8:172, 177; 9:117, 176, and see evolution, genetics, phylogeny
Teaching of statistics, 5:96; 6:85, and see statistics curricula
  aids, 2:97
  methods, 5:95
Tests, 3:3, 22, 40, 137; 4:214; 5:75, 84, 236; 6:14, 17, 100, 114, 179, 192, 237, 298, 400; 7:13; 8:11, 36, 275; 9:11, 45; 10:92, 167, 375, 479, and see acceptance area, analysis of variance, approximate significance

levels, chi square, conclusions, confidence limits, duo-trio test, F test, future action, goodness of fit, g statistics, Hotelling's T, inference, least significant difference, likelihood ratio test, nonparametric tests, nuisance parameters, null hypothesis, operating characteristic, ranks, rejection area, rejection of data, sign test, significance, significant cut, small samples, studentized extreme deviate, triangle test, t test
approximate, 1:26; 5:75, 84, 101, 211; 6:308
choice of, 7:51
combination of, 1:84; 3:125; 5:338; 9:76, 255; 10:433
combinatory, 10:311
composite, 5:75; 7:18, 73; 9:531; 10:410
compound symmetry, see symmetry tests
destructive, 6:229
distribution-free, see nonparametric
efficiency, 1:81; 6:167
exact, 3:19; 5:101; 6:299; 8:116; 9:305; 10:73, 420, 446. 516
interpretation of, 1:8, 26, 38, 70; 2:10, 14, 16, 27, 29, 38, 47, 56, 95, 107, 113, 123; 3:98; 4:264; 7:299, 312; 9:398; 10:417, 445
logic of, 1:47, 51; 2:29
metal, see psychological
multiple, 5:84; 9:262, 531; 10:168, 447
multivariate, 4:50; 5:78, 81; 6:259, 308, 374; 7:51; 9:20; 10:211
non-normal data, 2:73; 5:242; 6:191; 10:303
normality, 8:18
of significance of
correlation, 1:71; 2:27; 3:54; 4:163; 6:251
difference, 6:114
difference between adjusted means, 2:46, 54, 58; 4:244
difference between means, 1:74; 5:100; 8:33
difference between proportions, 2:95; 3:100
difference between regressions, 3:47
differential viability, 7:271
dominance deviation, 4:263
extreme deviate, 3:28; 9:84; 10:425
table, 9:89
fitted constants, 2:23
largest value, see extreme deviate
largest difference, 1:26; 5:99, 250; 7:299
linearity of regression, see linearity
mean difference, 2:95
multiple correlation, 2:27
non-additivity, 5:84
non-normality, 3:26
range, 1:26; 2:16
regression, 2:62; 4:160, 249; 6:282; 8:115, 384
regression intercept, 3:66
results suggested by data, 5:100; 6:166; 9:74, 534; 10:425
trend, see regression
one-sided, 1:60; 3:25, 119; 8:204, 391; 10:70, 543, and see two-tailed tests
power of, 1:49; 3:22; 6:179, 189; 8:222, 391; 9:75, 84; 10:142, 303, 417, 449, 479
psychological, 5:77, 78; 8:388; 9:265, 267
range, 9:262, and see tests using range
robust, 10:303
sensitivity, 6:229
sensitivity of, 2:96; 3:22; 5:176; 6:305; 10:417, 458
sequential, 6:61; 10:311
power of, 10:312
studentized, 5:101
successive, 5:100
symmetry, 5:81; 6:259
tables, 6:280
transformations, 5:284
two-tailed, 1:70; 2:56; 3:24; 4:264; 6:286; 7:19, 222; 8:36, 391, and see one-sided tests
using range, 5:105, 333
validity of, 2:18, 47; 8:170; 10:290, 401, 436, 479
variances heterogeneous, 6:14; 10:303, and see Behrens-Fisher test
z, see F test
normal deviate, 1:60; 6:412; 7:26; 10:429

random, 5:106
Tetrads, 4:11; 6:182; 7:60
Tetraspores, 7:118
Theory, see hypothesis, model
Threshold, 8:15, 175, 388, 389, 390; 9:263, 279; 10:185
Time response curve, 3:172; 4:166; 6:99, 174, 344; 7:120; 8:13, 307; 9:445; 10:531, and see dose response curve, extinction time
tables, 8:353
truncated, 7:120; 10:531
Time series, 1:16; 3:57; 5:243, 337; 6:176, 191, 283, 364, 437; 9:264, and see correlated errors, cycles, economics, harmonic analysis, moving averages, stochastic processes
Toxicology, 1:4, 42, 48, 80, 85; 2:57; 3:44, 172; 5:83; 6:298, 326; 7:120, 321; 8:13, 249, 311; 10:143, 155, 291, 310, and see bioassay, pharmacology
Transformations, 1:85; 2:76; 3:26, 39, 174; 4:91, 197, 211; 5:73, 86, 166, 237, 261, 282; 6:145, 180, 210, 343, 345, 437; 7:120, 212, 235; 8:249; 9:50, 191, 209, 427; 10:62, 130, 563, and see additivity, analysis of variance, bioassay, canonical analysis, logistic curve, models, variance of function
angular, 1:15, 43; 3:45, 62; 6:10, 180, 321; 8:2; 9:45, 467; 10:130, 447
table, 9:470
arc hyperbolic sine, 3:43; 9:384
arc sine, see angular transformation
by differencing, 6:364
choice of, 8:2; 10:564
effect of, 2:74; 3:174
empirical, 10:137
equivalent deviate, 9:467
inverse sine, see angular transformation
legit, 6:357; 10:131
table, 6:357
linear, 6:378
logarithmic, 1:11, 43; 2:74; 3:25, 39, 43, 195; 5:72, 264; 6:114, 146; 7:113, 174; 8:217; 9:290, 425; 10:196, 263, 412, 532, 564
log $(x + c)$, 2:74; 10:147, 564
logit, 1:39; 3:49, 62; 4:199; 6:321, 359, 432; 7:302, 327; 8:8; 9:471; 10:131, 148, 228, 310, 399, 443
loglog, 5:130; 10:131
table, 5:136
normal deviate, 3:49, 104; 9:32, 108
of percentages, 3:49
of ranks, 2:76; 9:32, 108; 10:392, 525
probit, 1:31, 39, 41, 42, 57; 3:41, 47, 62; 4:199; 5:83, 85, 86, 130; 6:225, 321, 359, 413, 424, 432; 7:302, 327; 8:2, 18, 96, 120, 179, 249, 310; 9:329, 468; 10:130, 173, 228, 306, 310, 314, 399, 443, 534, 556
slide rule, 1:46
tables, 5:85, 86; 6:426
reciprocal, 6:101; 10:532
rectangular, 9:471
selection-diffusion, see legit transformation
square-root, 1:43, 85; 2:74; 3:41, 91; 6:180; 7:174; 8:384; 10:135, 446
table of, 3:52
triangular, 6:186
z, 3:45, 98, 127; 8:2; 10:137
Tree crops, 6:164, 168; 8:89; 9:428, 429; 10:176, and see field experiments, long-term experiments, perennials
Trend, 1:38; 2:5, 61; 3:137, 154; 4:184; 5:41, 76; 6:229, 241, 282, 312, 218; 9:260, 304, 339, 429; 10:421, and see regression, time series
essential, 6:287
Triangle test, 4:80, 5:337; 6:102; 9:24, 43; 10:521
Truncation, see censored data, and under various distributions
artificial, 10:549
partial, 10:536
T test, 1:8, 26, 38, 59; 2:13, 27, 38, 96; 3:14, 24, 98, 101; 4:214, 272; 5:101, 124, 210, 337; 6:123, 191, 298; 7:52, 125, 299, 315; 8:5, 222, 384; 9:16, 32, 66, 109, 314, 531; 10:96, and see Behrens-Fisher test, , confidence limits, Hotelling's T
computation, 7:315
noncentral, 8:29
non-normal data, 6:191
one-sided, 3:25; 5:287

power of, 1:50; 7:312; 8:222
two-tailed, 3:24
Tuberculin, 3:173; 4:100; 5:300
Tuberculosis, 3:133; 4:204; 6:172; 10:195, 313
Twins, 8:187; 9:265
Type III distribution, see gamma distribution
Type VI distribution, 6:321
Unequal subclasses, 7:11, and see disproportionate subclasses
Uniformity data, 3:26; 4:100, 115; 5:144; 6:34; 7:168, 300, 310; 9:412, 429
U-shaped distribution, 2:88; 8:223
Variance, see Behrens-Fisher test, biological variation, coefficient of variation, covariance, *F* test, genetics, incomplete beta function, index of dispersion, mean square error, Sheppard's correction, significantly small *F*
about regression line, 2:105; 3:137; 5:69; 7:189, 260; 9:502
analysis of, 1:15, 16, 25, 26, 38, 42, 70, 77, 80, 85, 86; 2:7, 14, 16, 25, 26, 31, 36, 37, 56, 58, 76, 78, 84, 87, 89, 110, 118, 123; 3:1, 22, 49, 72, 91, 97, 137, 138, 195; 4:40, 55, 64, 101, 111, 132, 135, 173, 235, 245, 256, 267; 5:11, 68, 72, 84, 88, 99, 169, 174, 176, 192, 232, 250, 279; 6:8, 25, 99, 178, 259, 293, 295, 368; 7:3, 21, 37, 76, 85, 98, 125, 174, 290, 433; 8:57, 74, 86, 122, 169, 238, 265, 374, 382, 384; 9:14, 30, 45, 75, 131, 145, 163, 244, 263, 300, 337, 387, 422, 429, 434, 468, 501, 528; 10:62, 105, 130, 165, 167, 185, 196, 235, 259, 273, 283, 291, 300, 302, 308, 327, 401, 431, 490, 562, and see additivity, analysis of covariance, chi square, combination of data, components of variance, degrees of freedom, disproportionate subclasses, errors, fitting constants, *F* test, harmonic analysis, incomplete experiments, interaction, interblock error, interblock information, intrablock error, least significant difference, least squares, linear hypothesis, long-term experiments, main effect, missing values, models, multiple classification, multivariate analysis, orthogonal polynomials, pooling, regression, separation of variance, tests, transformations, uniformity data
computation of, 1:18, 85; 2:36, 38, 116; 3:6, 30, 47, 137; 4:65, 104, 173, 212; 5:5, 149, 234, 330; 6:35, 180; 7:124, 289; 8:78, 141, 370; 9:168, 308, 391; 10:76, 257, 333
incomplete data, 10:345
interpretation of, 2:37, 63; 5:84, 99; 6:293; 7:70; 8:384; 9:170, 389; 10:187
of variances, 7:113; 8:217
power of, 9:75
proportional subclasses, 7:70
approximate, 2:54; 3:167; 4:185; 9:60; 10:470
asymptotic, 7:10, 137; 8:24, 318; 9:182, 261, 486; 10:404, 548
biased, 8:347; 9:79
complex, 2:110; 3:64; 4:135
components, 1:18, 25, 43; 2:7, 15, 27, 34, 37, 77, 110; 3:4, 31, 75, 92, 98; 4:64, 107, 127, 135, 223, 234, 254; 5:72, 151, 189; 6:8, 25, 80, 346, 368; 7:1, 17, 36, 70, 75, 83, 97, 258, 287, 301; 8:58, 238, 265; 9:226, 262, 337; 10:57, 155, 165, 239, 304, 330, 408, and see components of covariance
computation of, 1:18; 5:192; 6:179; 9:227
confidence limits for, 6:136; 7:55, 80
interpretation of, 9:345
negative, 6:137; 7:4, 41, 89
variance of, 2:9; 7:55, 77
with covariate, 7:12
composite, see complex variance
conditional, 9:65
distribution of, 2:9; 7:80; 8:177
effective, 5:156
error, 1:42, 56; 2:3, 31; 3:23, 35, 75, 94; 4:156, 181, 212, 215; 5:5, 265; 6:25, 43; 7:3, 14, 58; 8:58, 93, 125,

170, 239; 9:132, 173, 256, 306, 473; 10:32, 504
estimated, 9:447
homogeneity of, 1:42, 63, 74; 3:12, 92, 195; 4:125; 5:263; 6:14, 259; 7:19, 103, 113, 189; 8:242; 9:11, 61, 210, 427, 468; 10:104, 244, 516, and see heterogeneity of error, scedasticity
multivariate analogue, 6:386
interblock, 2:58; 5:5, 151; 6:25
intrablock, 2:58; 5:5, 154; 6:25, 43
matrix, see covariance matrix
mean, 3:47; 4:71, 110; 5:156; 6:49; 8:132; 9:148, 221
minimum, see minimum variance estimate
negative, 5:72
of adjusted mean, 2:42, 53, 58, 118; 3:77; 4:71; 6:27; 7:292; 10:70
of asymptote, 7:259
of chi square, 10:462
of count, 8:59
of criterion, 6:259, 285
of dependent variable, 1:37
of difference, 2:113; 9:312
of difference of adjusted means, 4:71, 181, 244; 5:6, 86, 156, 159; 6:27, 47; 8:43, 78, 125; 9:131, 148; 10:71, 261, 277, 288
of difference of means, 2:36, 46; 6:248; 8:42; 9:533; 10:336
of estimate, 3:166; 4:106; 6:218, 401; 7:137, 156; 8:5, 27, 279, 333; 9:94, 181, 202, 293, 321, 400, 463, 486, 508; 10:178, 230, 404
of estimate of component of variance, 7:4, 87
of estimate of heritability, 10:312
of estimate of independent variable, 5:141
of extrapolation, 7:257
of function, 2:96, and see variance of linear function
of gene frequency, 1:27
of intercept, 3:61; 5:140; 7:261
of linear function, 3:11; 10:424, and see variance of function
of mean, 4:135; 5:191, 205; 7:86, 97, 312; 8:388; 9:59, 173; 10:105
partially weighted, 10:127
unweighted, 9:451
weighted, 9:59, 449; 10:126
of mean difference, 2:31
of mean score, 10:437
of median, 9:75
of product, 6:411; 7:301; 8:108
of proportion, 1:15; 6:391; 9:291; 10:480
of rank total, 3:119
of ratio, 5:335; 6:411
of regression coefficient, 1:37, 60; 2:34; 4:30; 5:140, 208; 7:65; 9:313
of regression estimate, 4:252, and see variance of adjusted mean
minimum, 9:332
of scores, 6:8
of sex ratio, 3:101
of total, 2:31; 8:112
of variance, 2:110; 3:45; 9:66, 267, 463
pure, see components of variance
ratio, 1:70; 2:28, 37, 55, and see *F* test
rectangular lattices, 5:85
regression on age, 9:267
related to mean, 2:3; 3:23, 32, 40, 91; 5:72, 301; 6:1, 100, 129, 318, 435; 9:45, 468; 10:117, 196, and see correlation of mean with variance
replacement, 7:56
sampling, 3:92, 100, 163; 4:226, 235; 5:174, 189, 340; 6:60; 8:287; 9:173, 176
separation of, 7:103, and see components of variance
subnormal, 3:175
test of, 10:421
theoretical, 9:471
Vector methods, see multivariate analysis
Viability, differential, 4:5; 5:305; 7:270
Virology, 1:43, 63; 2:34, 81; 4:205; 6:169; 7:320; 10:229, 511, and see half-leaf method, plant pathology
mixed infections, 2:82
Vital statistics, see actuarial statistics, birth rate, clinical statistics, death rate, demography, medicine, morbidity, mortality, sex ratio
Vitamins, 1:37; 6:294
A, 1:62, 3:173
ascorbic acid, 1:4; 2:78

B, 5:277; 8:371
D, 1:62; 2:34; 4:49; 5:277; 6:321; 7:215
E, 1:61
riboflavin, 2:78
thiamine, 2:78
Wear curve, 6:362, and see survival curve
Weber-Fechner law, 1:43
Weighted squares of means, method of, 2:10, 24; 4:244; 9:264
Weighting, 1:39, 41, 48, 49, 60; 2:12, 58 117; 3:11, 29, 41, 47, 62, 78, 99, 102; 4:66, 119: 5:5, 86, 130, 157, 168, 303; 6:10, 26, 43, 224, 319, 360; 7:64, 190, 289, 302; 8:8, 93, 115, 133; 9:34, 59, 91, 117, 267, 291, 427, 468, 495; 10:101, 133, 155, 335, 412, 431, 443, 546, and see semiweights
efficiency of, 9:447
partial, 10:119
Wood decay, 6:296
Wool sampling, 7:83
Wound healing, 4:51; 8:224, and see growth
X-rays, see radiology
Yates' correction, see continuity correction
Youden squares, 2:36, 83; 7:117; 10:155, 251, 507
Zoology, 3:59; 7:300; 10:162
Z test, see *F* test